目　录

第一篇　生物大分子X射线晶体结构测定的基本原理 …… (1)
第一章　生物大分子的晶体培养 …… (2)
1.1　影响蛋白质溶解性的因素 …… (3)
1.2　影响晶核形成和晶体生长的因素 …… (4)
1.3　结晶技术、 …… (5)
第二章　晶体结构的共同特征 …… (9)
2.1　晶体的空间格子、晶胞和晶面指标 …… (9)
2.2　晶体的对称性、点群、晶系、空间群 …… (14)
第三章　晶体的X射线衍射花样 …… (25)
3.1　衍射线的分布 …… (25)
3.2　衍射线的强度 …… (35)
第四章　记录衍射花样的方法 …… (42)
4.1　X射线的产生和一般性质 …… (42)
4.2　蛋白质晶体的挑选和安装 …… (43)
4.3　回摆照相法 …… (44)
4.4　旋进照相法 …… (45)
4.5　魏森堡照相法 …… (51)
4.6　四圆衍射仪的基本原理 …… (54)
4.7　生物纤维状物质的X射线衍射 …… (56)
第五章　晶胞中原子位置的测定 …… (62)
5.1　晶体的电子密度分布函数 …… (62)
5.2　电子密度图 …… (63)
5.3　电子密度图的解释与原子在晶胞中的位置 …… (65)
第六章　解决相角问题的方法 …… (69)
6.1　帕特逊(Patterson)函数法 …… (69)
6.2　同晶置换法 …… (71)
6.3　反常散射法 …… (75)
6.4　分子置换法 …… (77)
第二篇　生物大分子的立体结构 …… (85)
第七章　蛋白质的立体结构 …… (86)
7.1　氨基酸和多肽链的立体结构和构象角的定义 …… (86)
7.2　各种氨基酸对蛋白质三维结构的影响 …… (91)
7.3　纤维状蛋白质的立体结构 …… (93)
7.4　球状蛋白质的晶体结构 …… (97)
7.5　膜蛋白的晶体生长和晶体结构 …… (136)
第八章　核酸的立体结构 …… (146)

8.1 核酸的结构和构象 …… (146)
8.2 脱氧核糖核酸(DNA)的立体结构 …… (147)
8.3 核糖核酸(RNA)的晶体结构 …… (160)
第九章 病毒的立体结构 …… (166)
9.1 病毒的形态学和整体结构 …… (166)
9.2 病毒的结晶和样品制备的一般原理 …… (168)
9.3 病毒的X射线衍射分析研究简述 …… (170)
9.4 具有螺旋对称性的病毒 …… (170)
9.5 球形病毒 …… (178)
第十章 核小体和染色体的立体结构 …… (190)
10.1 核小体核心颗粒的结晶和单晶结构分析 …… (191)
10.2 核小体核心颗粒的立体结构 …… (193)
10.3 染色质的X射线纤维衍射分析和染色质超级结构的螺线管模型 …… (196)
第十一章 多糖的立体结构 …… (199)
11.1 多糖的构象 …… (199)
11.2 多糖的X射线衍射分析方法 …… (201)
11.3 多糖的立体结构 …… (204)
参考文献 …… (209)

第一篇　生物大分子X射线晶体结构测定的基本原理

生物大分子X射线晶体学是应用X射线衍射方法研究生物大分子的晶体，揭示分子结构与功能关系的科学。因此为了掌握生物大分子X射线晶体学的基本原理，必须首先对X射线衍射法有一个基本的了解。

X射线是一种电磁波，能直线传播，能使胶片感光，这是大家共知的性质。在晶体结构分析中应用的X射线是波长在0.1nm左右的电磁波，由于这一波长与晶体内原子间的距离具有同一数量级以及晶体结构内分子排列的规则性，当X射线入射到晶体上时，晶体中的每一个原子都发射出次生的X射线，并相互干涉，形成一个衍射花样。如果将这个过程与光学显微镜成像相比较，两者有相似之处。在显微镜下，一束平行的可见光入射到一个物体上，物体散射入射光，物镜会聚散射光而成像。对于X射线来说，由于还没有一种材料可作为聚焦镜，因此不可能直接会聚散射光而形成物像，看到晶体内部的结构。但是晶体的衍射花样与晶体内部的结构有一定的关系，即衍射花样内衍射点的排列方式、点间距离的大小与晶体内生物大分子的排列方式和重复周期大小有关，而衍射点的强度分布与生物大分子结构本身的特点有关。因此，从原则上讲我们可以通过分析衍射点的排列方式和测量点间距离的大小来推算分子在晶体结构中的排列方式和重复周期的大小，以及通过测量衍射点的强度，应用一系列数学方法，借助电子计算机可测定分子内每个原子在空间的坐标，从而测定整个分子的结构和晶体结构。

生物大分子X射线晶体结构的测定工作，在实践中要经过一个较复杂、长期的过程，大体上可以分成彼此相互联系和影响的若干步骤：(1)培养大的、质量好的晶体，以及进行初步的X射线衍射分析；(2)重原子衍生物的制备；(3)衍射数据的测量和处理；(4)相位的计算；(5)电子密度图的计算和解释以及分子模型的修正。在上述几个方面，晶体学家已积累了许多成功的经验，并有专门的论著。

为了进一步理解X射线衍射法测定晶体结构的基本原理，本篇将首先讨论生物大分子的晶体培养和介绍与晶体产生衍射花样有关的、最基本的晶体结构特征；其次，讨论衍射花样中衍射点的分布与晶体内分子重复的方式和重复周期大小的关系，衍射强度与晶体内分子空间结构的关系；第三，讨论X射线的产生和性质以及记录衍射花样的一些方法和原理，以帮助读者进一步理解衍射花样的产生和晶体结构的关系，并对生物样品的纤维衍射图的特征，作简要的说明；第四，由衍射花样测定了晶体的晶胞大小、形状，以及测量了衍射强度，并引出结构振幅和解决相角问题以后，将应用电子密度函数计算晶体的电子密度图，最后通过解释电子密度图和不断修正分子模型，得到晶胞中每个原子的坐标。

110740

第一章　生物大分子的晶体培养

要进行X射线晶体结构分析，首先要得到适合于结构分析的晶体。这里所谓"适合于"包括两层意思：第一，晶体内部结构要具有有序性，是单晶，不是孪晶，否则无法得到具有结构本身特点的衍射花样；第二，晶体要有一定的大小和形状。因为晶体衍射线的强度大体上正比于晶体的体积，而反比于分子量的大小。一般讲分子量为50000左右的蛋白质分子，需要0.3mm^{3}或者更大的晶体，才有可能作高分辨率的结构分析。对于分子量更大的蛋白质分子，那就需要更大的晶体。为了满足上述要求，首先要使生物大分子结晶，然后设法长大。

应用X射线衍射法研究生物大分子的晶体结构，当前最主要是蛋白质分子和核酸分子以及它们的复合物，而蛋白质和核酸分子两者在表面电荷、外形等方面虽有所不同，在结晶条件上亦有差异，但是基本规律、方法还是一致的。所以，在此主要以蛋白质分子为例，来讨论生物大分子晶体培养的一般规律和方法。

蛋白质结晶过程像其他小分子物质一样，是一个有序化过程，即在溶液中处于随机状态的分子转变成有规则排列状态的固体。一般认为要使这种有序化过程开始必须要形成一定大小的晶核，并使分子不断地结合到形成的晶核上。而一个蛋白质溶液能开始形成晶核，就必须使溶液达到过饱和，并保持一定的条件，使溶液中的分子失去自由运动的能量（平移、旋转等）而结合到晶核上，形成新的稳定的化学键（次级键），使整个体系能量降低而形成晶体。

蛋白质、核酸等生物大分子的结晶比一般小分子化合物的结晶要困难得多，这是由于这些大分子的分子量很高，几何形状较复杂，表面带多种电荷，因此分子间相互作用或相互结合的点很多，而可能形成有序排列的关键结合点和几何匹配位置又很少，在外界条件（如pH、温度、不同溶剂等）的影响下，分子构象容易产生某些变化；同时，在大分子结晶时，又必须保持在水合状态，或者在生理pH和温度条件下。因此，一般小分子结晶的方法，都不适合应用于大分子的晶体生长，从而使得生物大分子的晶体培养工作相当困难。不过在蛋白质和核酸的晶体培养方面，经过长期的努力已积累了许多成功的经验，对蛋白质的溶解、结晶和生长等方面也进行了许多理论和实验研究。长期以来很难培养出单晶的膜蛋白质，也已培养出了适合于X射线衍射分析的晶体。因此，在某些方面已有可能根据生物学意义来选择蛋白质和核酸的X射线晶体结构分析的课题，改变了以往只能有什么晶体就分析什么晶体的状态。但是从总的方面来说，生物大分子的晶体培养仍然是摸索性和经验性的，在某种程度上还是靠机遇。因而许多重大的生物学课题由于得不到适合于分析的晶体而长期不能进行。晶体培养已成为当前阻碍提高晶体结构分析速度的关键性问题。为此人们正在作进一步的努力。在物理学方面，正在研究生物大分子的结晶、生长的复杂机理；在自动化方面，已设计和使用了由计算机控制的蛋白质结晶系统；在微重力条件下的晶体生长研究也已引起了不少科学家的兴趣和重视。

要培养蛋白质或核酸的晶体最重要的是分离、纯化，获得均一的样品，一般要求电泳纯。然后使溶液达到过饱和，在这个过饱和状态下蛋白质才可能开始结晶。蛋白质溶液的饱和度是随溶解性的变化而变化，而溶解性是与蛋白质的浓度、离子强度、温度、pH、加入的有机溶剂以及与蛋白质结合的相反离子等等有关。

1.1 影响蛋白质溶解性的因素

蛋白质分子是具有一定几何外形的结构，它的带电的、极性的基团分布在分子的表面，它的溶解性质可以像任何其他类型的水溶性分子一样来考虑。在溶液中的水分子与蛋白质表面的带电基团和极性基团相互作用，从而在分子周围形成一个水层。由于水的高介电常数，使得带电的分子不能相互结合而沉淀，从而使蛋白质分子溶解于水溶液之中。

1.1.1 离子强度

蛋白质分子可简单地看成是一个大的多价离子，它的溶解性依赖于溶液的离子强度。而离子强度 $\mu=\frac{1}{2}\sum C_iZ_i$，$C_i$ 为离子浓度，Z_i 为离子价。当把中性盐加入到蛋白质溶液中时，在低浓度时离子与蛋白质分子表面电荷相互作用形成离子层，这种离子层更好地与水作用而增加了蛋白质的溶解性。这种现象称为盐溶。而当盐浓度增加到一定数量时，离子之间，蛋白质与离子间相互争夺水分子，结果导致了蛋白质分子周围的水化层的破坏而降低了蛋白质的溶解性。此现象称为盐析。

蛋白质分子比一般离子大得多，它的表面极性基团、带电基团相当复杂，表面的有序溶剂分子，以及蛋白质分子的固体状态等都会影响蛋白的溶解性。但是根据盐溶、盐析现象，可以改变蛋白质的溶解性以达到控制蛋白质溶液的饱和度，引导蛋白质结晶。

不同离子对蛋白质的溶解性的影响是不同的。小的高电荷离子比大的低电荷的离子在盐析方面影响大，如氯化钾对蛋白质的盐析影响较小。一般讲磷酸钾、硫酸钠、硫酸铵、柠檬酸钠、硫酸镁对蛋白质溶解性的影响按顺序减小。总之，盐对蛋白质溶解性的影响正比于离子所带的电荷和浓度(图 1.1)。

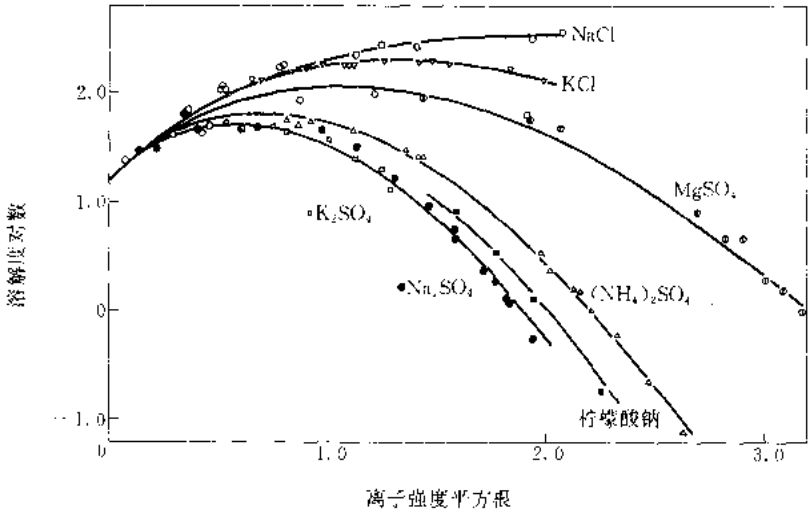


图 1.1 在不同电解质中一氧化碳血红蛋白的溶解性(25℃)

1.1.2 pH 和相反离子

把蛋白质分子看成具有确定价的带电荷的物质，显然是过于简单，是不合适的，因为可以

通过改变 pH 或加入特定的结合离子(相反离子)到蛋白质的极性基团上,从而改变蛋白质分子的带电性。一般来说,蛋白质分子带的电荷越高,它的溶解性越大,当纯电荷为零时它的溶解性最小。因此可以通过改变 pH 或加入相反离子来改变蛋白质的溶解性。pH 的改变对蛋白质溶解性的影响见图 1.2。

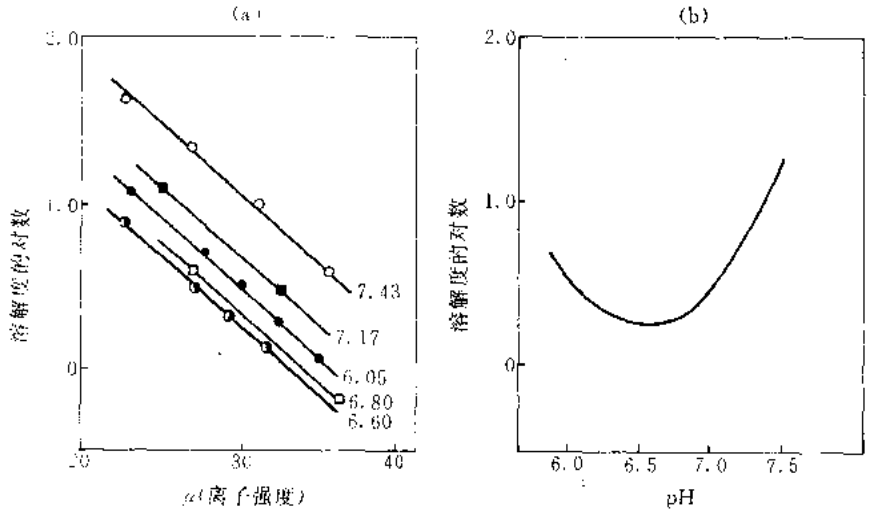


图 1.2　在不同 pH 的浓磷酸缓冲液条件下血红蛋白的溶解性

1.1.3　温度

温度对蛋白质溶解性的影响较复杂。一般讲,在低离子强度下通常是随温度的增加而增加,而在高离子强度下蛋白质的溶解性随温度的增加而降低。

1.1.4　有机溶剂

有机溶剂是影响蛋白质溶解性的重要因素。在一定范围内有机溶剂亦是通过争夺蛋白质分子表面的水分子,以及降低介质的介电常数,减小介质的有效静电屏蔽,增加蛋白质分子间的静电相互作用,从而降低蛋白质的溶解性。在使用有机溶剂时,要尽可能在低温条件下操作,以免产生蛋白质变性。

1.2　影响晶核形成和晶体生长的因素

通过改变离子强度、pH、相反离子类型、温度和有机溶剂浓度等因素可以减少蛋白质的溶解度,最后使溶液达到过饱和。而过饱和是一种不稳定的状态,在这种状态下产生晶核,其中一些晶核生长成大的晶体。但是,不稳定状态向稳定状态的转化可以加以控制,从而达到产生较少的晶核以及生长出大的晶体。

在生物大分子晶体培养中最常用的沉淀剂可归纳为三大类:盐类、有机溶剂及聚乙二醇类。最常用的盐类有硫酸铵或硫酸钠、柠檬酸钠或柠檬酸铵、氯化钠或氯化钾和氯化铵、硫酸镁、氯化钙、硝酸铵、甲酸钠等等。常用的有机溶剂有乙醇、异丙醇、丙酮、二甲基-2,4-戊二醇(MPD)等。聚乙二醇(PEG)最常用的平均分子量为 2000—6000,但是从 400 至 20000 大小的分子量都成功地用于蛋白质结晶。聚乙二醇分子量的大小对促使蛋白质结晶有些影响,但是,尚未发现有明显的、确定的关系。用于结晶的聚乙二醇的浓度,一般在 4%至 18%范围以内。聚

乙二醇的作用主要是改变原来水的结构，而与水形成复杂的网状结构，从而促使溶剂排斥蛋白质分子而减少溶解性，最后使蛋白质结晶。聚乙二醇是目前较为广泛使用的沉淀剂之一。

由于蛋白质分子溶液达到饱和、过饱和状态的情况较复杂，在平衡过程中有离子和分子间的相互作用，因此对于如何产生晶核和长成大的晶体了解得还很不够。根据对小分子物质结晶的研究，晶体的形成取决于一个一定大小的晶核的形成，如果产生的晶核太小就可能溶解。一定大小的晶核可能包含 10 至 200 个分子，形成的时间随条件而变化，成核的速度明显地随过饱和度而变化，溶液达到某一过饱和度时，成核速度就迅速增加。为了减少成核的数量，从而控制晶体的数量，必须尽可能地使溶液慢慢地达到过饱和度，以及使过饱和度尽量地低。因此加沉淀剂时应该尽可能地慢，以免局部产生过饱和。

在溶液中某些外来的微粒可能作为成核中心，而使晶核更多，因此应该尽量避免有尘粒，或离心除去；在实验过程中要尽可能避免产生小的气泡。在使用有机溶剂作沉淀剂时要特别小心，因为稍有不慎很可能引起蛋白质变性，产生不必要的成核中心；通常蛋白质应该很纯，蛋白质溶液不要保存太久，以免蛋白质变性。因此在开始培养晶体前，应首先将准备结晶的蛋白质溶液高速度离心。

蛋白质的浓度对晶体的生长亦是一个重要的因素。大的晶体往往在高浓度、低过饱和度情况下产生，这是由于高浓度蛋白质溶液提供了足够量的分子，满足了晶体生长的需要。如果可能向结晶的溶液中引进一些晶种，有利于减少成核，生长出大的晶体。

1.3 结 晶 技 术

要使生物大分子结晶和生长出大的晶体，最关键的是控制过饱和度的量和速度，过饱和度要低，而速度要尽可能地慢。为达到上述目的，下面介绍一些在实践中应用较成功的培养晶体的技术。

1.3.1 微量蒸汽扩散法

小分子结晶一般用蒸发溶剂使溶液浓度增加的办法来结晶，但是这种方法不适合于蛋白质的结晶，因为蒸发的程度很难控制，而往往使蛋白质溶液内的盐首先结晶，或者使整个蛋白质水合体系失水而使蛋白质变性。为克服这一缺点，根据蒸发扩散的原理，设计了悬滴法和坐滴法。这些方法最关键的地方就是使某种蛋白质结晶所需盐的浓度与略低于这种盐浓度的蛋白质溶液在一个封闭体系内蒸发扩散，最后达到平衡而使蛋白质溶液内盐浓度增加，蛋白质溶解性降低，达到过饱和而结晶。

1.3.1.1 悬滴法

这种方法最适用于微量的筛选结晶条件。每个蛋白质溶液的样品只需 5μl 或者 10μl，平衡液只需 1ml。因此每次实验可根据需要设计，例如可先配制一系列不同浓度的沉淀剂溶液，作为平衡液，然后转移到各个封闭小室内，与含有低于平衡液 50%的沉淀剂的蛋白质悬滴液扩散平衡(图 1.3)。最后根据实验结果筛选出一种沉淀剂浓度作为最佳结晶条件。具体实验可采用 16 孔的塑料组织培养盒进行，每个孔内加入一定量的平衡液。每一个蛋白质母液悬滴加在预先硅化好的盖玻片上，然后把盖玻片倒盖在小池上。为防止蒸发扩散时漏汽，必须在盖玻片与池边缘间加少量凡士林密封。

1.3.1.2　坐滴法

原理与悬滴法相同，不同之处仅为液滴体积增大，如 50μl 或更大(图 1.4)。这种方法较易生长出大的晶体，重复性较好，但是蛋白质用量较大，适合于已大体知道结晶条件，或者有足够量的样品可供使用。其缺点是晶体有时贴在容器底部而不易取出，导致损坏晶体。

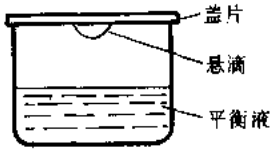


图 1.3　示意悬滴法装置

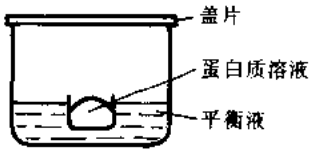


图 1.4　示意坐滴法

1.3.2　平衡透析法

利用半透膜允许小分子透过而不允许大分子透过的性质，来调节蛋白质溶液的离子强度或 pH 值，使蛋白质溶液慢慢地接近过饱和度。这种方法的优点是：通过有控制的扩散，改变结晶的条件，从而慢慢地到达晶核形成点。晶体的生长亦可不断地调整，已经产生的沉淀或者微晶可通过外部条件的改变重新溶解。这种方法可以在低离子强度条件下应用传统的盐析方法结晶，它既适用于大批量的结晶，亦可应用于微量的结晶。

1.3.2.1　微量扩散小室法

较普遍使用的微量扩散小室是用有机玻璃加工而成的，小室直径 2—3mm，高 4—5mm。蛋白质母液加在小室内，上端封一半透膜，然后把小室放到相应的平衡透析液内。蛋白质母液内的条件通过半透膜与外部溶液平衡透析而改变(图 1.5)。

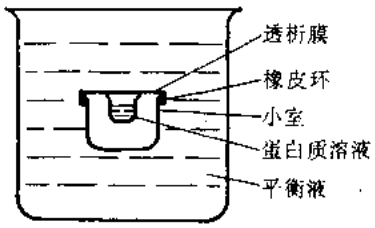


图 1.5　示意透析法装置

1.3.2.2　梯度透析法

为更慢地控制扩散速度，减少成核的数量可用具有双层透析膜的梯度透析装置，如图 1.6 所示。第一级透析装置可以用微量扩散小室，而第二级透析装置用体积较大的容器。透析液的体积和浓度根据实验设计要求逐级增大和提高，第一级溶液的浓度低于第二级溶液的浓度。

1.3.3　一批结晶法

这种方法的要点是控制所加沉淀剂的量而使蛋白质溶液逐步地达到低过饱和度。当溶液远离饱和状态时所加沉淀剂量的多少对达到低过饱和状态影响不大，但到临界于过饱和状态时，这个量的控制就较困难。因为完全可能在略为过量时，就使溶液达到过饱和而大量形成晶核。为了克服这种困难以及筛选某一种合适的条件，可采取配制一系列蛋白质浓度递减和沉淀剂浓度递增的溶液，观察在哪一种条件下产生较好的晶体，例如首先配小量的(300μl)低离子强度的高浓度蛋白质母液，然后从上述母液中取出 1/10 量(30μl)放在一个试管内，再向上述母液中加入 1/10 量(30μl)的溶剂至 300μl。这样蛋白质浓度就降低了 1/10 量。重复上述步骤，最后配得了一系列浓度递减的蛋白质溶液。然后对每一个 30μl 溶液按下述方法，配制成含有递增的沉

淀剂浓度的溶液。从 30μl 溶液中取出 3μl，放在一个小管内，然后再向溶液加入 3μl 沉淀剂至 30μl。然后再取出 3μl 溶液，再加入 3μl 沉淀剂。以此重复到所需的浓度条件为止。

经过一段时间的放置和观察，根据样品溶液产生沉淀或晶体的情况，就可得到有关样品溶解性与晶体形成之间关系的结果。

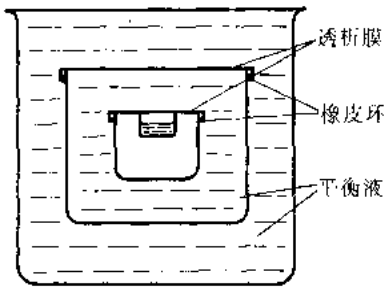


图 1.6　示意梯度透析法装置

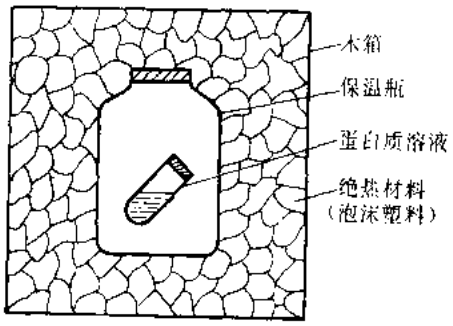


图 1.7　示意温度梯度法装置

1.3.4　利用溶解性的温度梯度结晶法

某些蛋白质在较高温度下，其溶解性高于室温下的溶解性，因此可以在较高温度下配制远离过饱和度的溶液，然后有控制地降低温度，使之慢慢地接近于过饱和度。例如在较高的温度下(40℃)配制低离子强度的蛋白质溶液，放置在保温瓶内，然后再放置在隔热的箱子内，让箱子慢慢地降低温度(图 1.7)。随着温度的降低，溶液达到过饱和度而结晶。

1.3.5　晶种接种或重复接种法

当某种蛋白质在应用各种微量结晶法试验后，仅能得到小晶体而得不到足够大的适合于分析用的晶体时，就可试验用接种或重复接种方法来培养晶体。这种方法是把已得到的小晶体经选择、清洗后接入到新配制的结晶母液中，作晶体生长的晶核，然后采用上面介绍的任何一种晶体培养方法继续设法使晶种长大成适合于分析用的大晶体。

晶种的来源：(1)由微量法得到的小晶体(0.01—0.1mm)，或者由其他方法得到的更小的微晶。(2)由簇生的针状的或片状的晶体，经精心分离、切断、挑选的单晶片段。晶体的选择，最好在偏光显微镜下进行，这样容易选取外形规整、无裂纹、又不是孪生的小晶体或晶体片段为晶种。

晶种的转移：选好的晶种，在接种前需在一定浓度的沉淀剂缓冲溶液中清洗，去掉晶种表面附着的晶体碎片和变性蛋白，并使晶体表面新鲜。沉淀剂浓度一般略高于原晶种生长时的浓度，以免在清洗过程中溶解。经多次清洗后的晶种，借助玻璃毛细吸管(图 1.8)或者用一端剪成铲状的动物毛发，小心地转移到新配制的晶体生长母液中。在转移过程中一定要保证晶种的完整，同时要注意尽可能地减少把清洗晶种的溶液带入新鲜母液，以防改变母液浓度。母液中蛋白质和沉淀剂浓度应根据原晶体生长时的条件，经试验决定。一般为蛋白质浓度略低于原晶种生长时的浓度，而沉淀剂浓度略高于原晶种生长时的浓度。这样既能保证晶种移入后不溶解，又有利于晶种的生长。

晶种的生长：已转移到母液中的晶种，可采用前面已介绍的任何一种微量晶体培养法促使

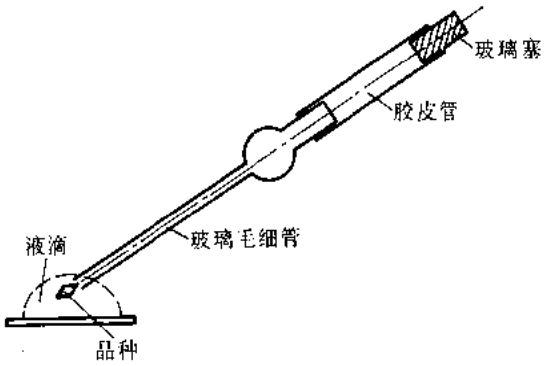


图 1.8　晶种的转移

晶种继续生长、增大。如果经一次接种后，尚不能达到足够大，可将该晶体再作为晶种，并按上述步骤再清洗、转移、生长。如还不能满足要求，可再重复接种，直至获得符合要求的单晶为止[10][13]。

第二章　晶体结构的共同特征

生物大分子如蛋白质、核酸，分子量很高，结构复杂，而且具有多种多样的生物功能，与一般的无机、有机分子有本质的差别。但是，当这些物质在一定条件下形成晶体时，它们都遵循着最普遍、最共同的规律，即整个晶体结构保持能量最低状态。由此，晶体结构内部分子（原子或离子）的排列必然具有规则性和最密堆积性。这种排列上的规则性和密堆积性集中反映在晶体结构都具有周期性和对称性这样的共同特征上。正是这些共同特征决定了晶体能作为光栅而使 X 射线产生衍射，以及晶体具有一定的熔点和各向异性等物理、化学特性。在这里主要讨论与 X 射线衍射法有关的共同特征。

2.1　晶体的空间格子、晶胞和晶面指标

2.1.1　晶体的空间格子

晶体是在自然界或实验室条件下，生成多面体形状的固体。虽然在某些生长条件下晶体的多面体外形可能不明显，但是天然的多面体外形是晶体外部最明显的共同特征。例如日常在实验室看到的氯化钠、硫酸铵等晶体，以及胰蛋白酶、胰岛素等生物大分子晶体，虽然化学组成和性质各不相同，但都具有规则的多面体外形。这一外部的共同特征反映了各种物质的晶体具有内部结构上的共同特征。

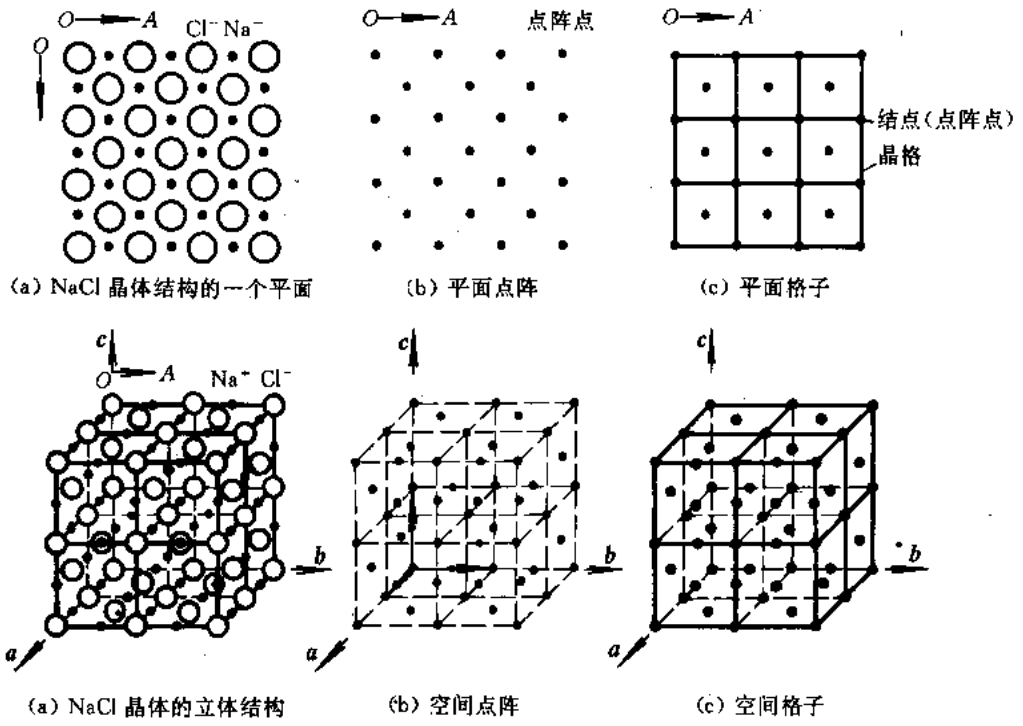
2.1.1.1　晶体结构的周期性

晶体结构内部最重要的共同特征是分子（原子、离子）在空间排列上的周期性，例如氯化钠（NaCl）晶体，Na^+和 Cl^-离子按一定周期在空间排列，形成三维的周期性结构。图 2.1(a)示出了这种周期性排列的一个平面和立体结构，其中 Cl^-和 Na^+都按 OA 周期重复排列。同样亦可以把 Cl^-和 Na^+看成是重复的基本单位，称为结构基元，重复的周期亦为 OA。又如 NH_4Cl 晶体，NH_4^+ 和 Cl^-离子亦按一定周期在空间排列，形成三维的周期性结构，图 2.2(a)示出了一个平面和立体结构，其中 Cl^-和 NH_4^+ 都按 OB 周期重复排列。NH_4^+ 和 Cl^-为结构基元，重复周期亦为 OB。当然 NaCl 和 NH_4Cl 晶体结构的周期大小不可能相等，但是原子排列的周期性是它们的共同特征。对于生物大分子来说，分子的结构虽然很复杂，但在晶体结构中，分子仍然按一定周期排列，即它们作为一个结构基元按一定周期排列。不过，由于生物大分子的分子量很高，体积很大，因而重复的周期要比一般小分子大得多。

在实际晶体里，这种重复的周期可以近似地看成是无限的，因为在一般小分子晶体中重复周期为 1nm 左右，而对边长为 1mm 的晶体来说，周期数可达 10^6 左右；蛋白质晶体的重复周期较大，假如为 10nm 左右，那末边长为 1mm 的晶体，周期数亦可达到 10^5 左右。

2.1.1.2　晶体的周期性与空间格子

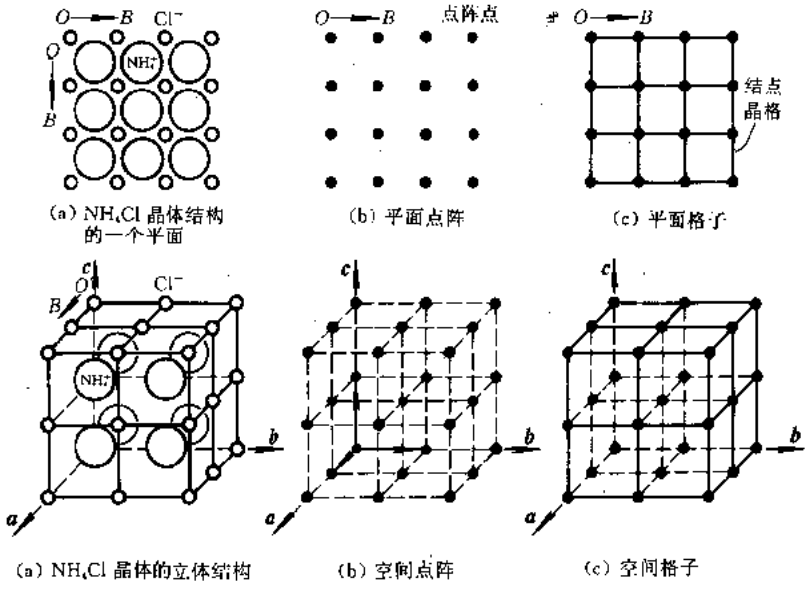
在晶体学中为了充分反映晶体结构的周期性特征，从实际晶体中抽象出了空间格子的概念。按照晶体结构的周期性，在一定的地方用一个点代表晶体结构的基本重复单位，就可抽象出在三维空间分布的一组点。例如在 NaCl 晶体中，用一个点代表基本重复单位（Na^+和 Cl^-），



(a) NaCl 晶体结构的一个平面　(b) 平面点阵　(c) 平面格子

(a) NaCl 晶体的立体结构　(b) 空间点阵　(c) 空间格子

图 2.1　示意 NaCl 晶体结构与点阵和空间格子间的关系



(a) NH_4Cl 晶体结构的一个平面　(b) 平面点阵　(c) 平面格子

(a) NH_4Cl 晶体的立体结构　(b) 空间点阵　(c) 空间格子

图 2.2　示意 NH_4Cl 晶体结构与点阵和空间格子间的关系

就可抽象出上述的一组点。图 2.1(b)为点在一个平面和三维空间中的分布。由于晶体结构的周期性是无限的，因此这样抽象出来的点在空间分布也是无限的，每个点的周围环境、相对的

定向是完全相同的。我们称这样的一组点为空间点阵。如果把空间点阵按确定的平行六面体单位划分，就构成空间格子，简称晶格，图 2.1(c)示出了 NaCl 晶体的一个晶格面和空间格子。同理，可从 NH_4Cl 晶体结构中抽象出三维空间分布的点阵、晶格[图 2.2(b)和(c)]。

一定的空间点阵按平行六面体单位划分的可能的方式是无限多的，如图 2.3 所示。但在晶体学上要求划分的平行六面体单位具有最高的对称性和最小的体积。因此对某一特定的晶体，它的平行六面体单位总是确定的。对各种不同物质的晶体，或者同一种物质的不同的晶型，它的平行六面体单位形状可以相同，亦可不同。每一个平行六面体单位的三个边长可以不相等，三个边相互间可以垂直，亦可不垂直。从各种物质的晶体结构中抽象出来的晶格的平行六面体单位共有 7 种类型：立方单位、三方单位、六方单位、四方单位、正交单位、单斜单位和三斜单位。[1][2][3]

2.1.1.3　14 种晶格单位型式

在每一种晶格类型中，根据不同的点阵点的分布情况，又可分成不同的晶格型式。例如前面已提到的 NaCl 和 NH_4Cl 晶体都是立方格子(图 2.1，图 2.2)，在单位格子的每一个顶角上都有一个点阵点，但在 NaCl 的每一个单位格子的中心还有一个点阵点。显然这两种单位格子具有不同的型式。另外还有一种立方格子型式，这种格子除在每个顶角上存在点阵点以外，在格子中心还有一个点阵点。这三种立方格子都符合空间格子的条件，只是点阵点的分布情况不同。仅在单位格子顶角上存在点阵点的称简单格子，其他两种型式称复格子。按照类似的方法分析其余晶格类型，总共可得到 14 种晶格型式。它们的名称、符号见图 2.4。图中三方(R)是三方晶格所特有的一种简单格子型式，R 表示菱面体 Rhombohedron。在《晶体学国际表》中将三方和六方晶格均按六方单位格子表达型式，得到简单六方和复六方两部分型式。[2][4]

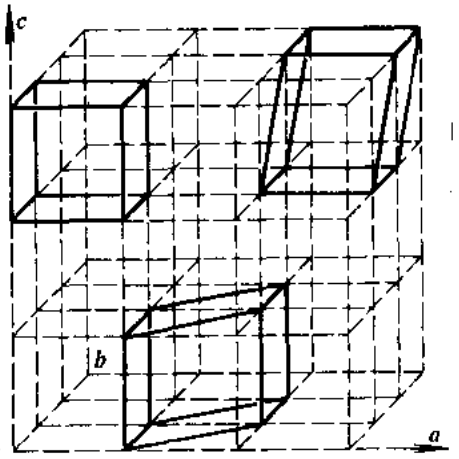


图 2.3　空间格子单位划分的多种类型

2.1.1.4　空间格子、基元与晶体结构

空间格子(点阵)是从实际晶体中抽象出来的概念，它最本质地反映了晶体中分子(原子或离子)在三维空间排列的周期性特征。空间格子结构是一切晶体结构的共同特征。任何晶体都可看成是由结构基元位于空间格子结点上所构成的固体。不同的晶体具有不同的空间格子类型和型式以及具有不同的结构基元。例如 NaCl 晶体是立方面心空间格子，结构基元是 Na^+ 和 Cl^- 离子；NH_4Cl 晶体是简单立方晶格，结构基元是 NH_4^+ 和 Cl^-；马一氧化碳血红蛋白晶体是单斜晶格，结构基元为 2 个血红蛋白分子。图 2.5 示出了上述血红蛋白晶体的晶格、基元与晶体结构三者的关系。

上面讨论了一切晶体结构具有的最基本的共性，但是生物大分子晶体尚有它的特殊性。它的空间格子是开放的，即在晶体结构内部除大分子以外，尚有大量的溶剂分子，形成许多小通道，可以同外界进行离子或小分子的交换。在今后晶体安装、数据收集等实验中必须充分注意到这一点。

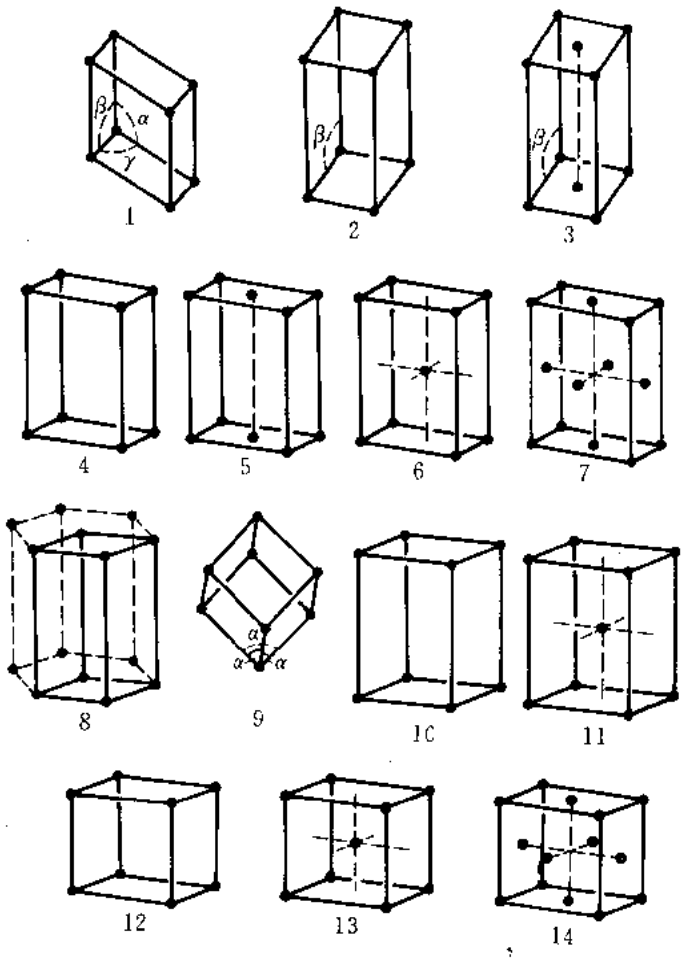


图 2.4　14 种晶格型式

1. 三斜(P)；2. 简单单斜(P)；3. 底心单斜(C)；4. 简单正交(P)；5. 底心正交(C)；6. 体心正交(I)；7. 面心正交(F)；8. 六方(P)；9. 三方(R)；10. 简单四方(P)；11. 体心四方(I)；12. 简单立方(P)；13. 体心立方(I)；14. 面心立方(F)

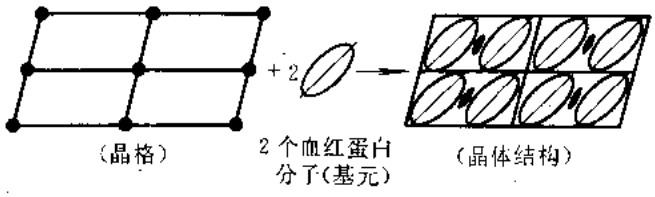


图 2.5　示意晶格、基元和晶体结构的关系

2.1.2　晶体的晶胞

晶体结构可以按晶格划分成一个个大小和形状完全相同的平行六面体单位。每个单位含

有相同数量的原子(分子),构成一个完整的结构单位,对整个晶体具有代表性。这样的平行六面体单位称为晶胞。整个晶体可以看成是由无数个晶胞在三维空间中有规律地、重复排列而成。因此在晶体结构测定中只要知道了晶胞的大小、形状以及原子在晶胞中的分布,就知道了整个晶体结构。

晶胞由 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 三个向量规定,晶胞的大小和形状由三个边长 a,b,c 和三个边之间的夹角 α,β,γ 表示。晶胞的 a,b,c 和 α,β,γ 称为晶胞参数。晶胞中任意一个原子的位置用分数坐标 x, y,z 表示,例如 NaCl 晶体的晶胞,$\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 三个向量相互垂直,轴长 $a=b=c$,夹角 $\alpha=\beta=\gamma=90°$。图 2.6 示出了 NaCl 晶体的晶胞参数及一个平面上原子的坐标。

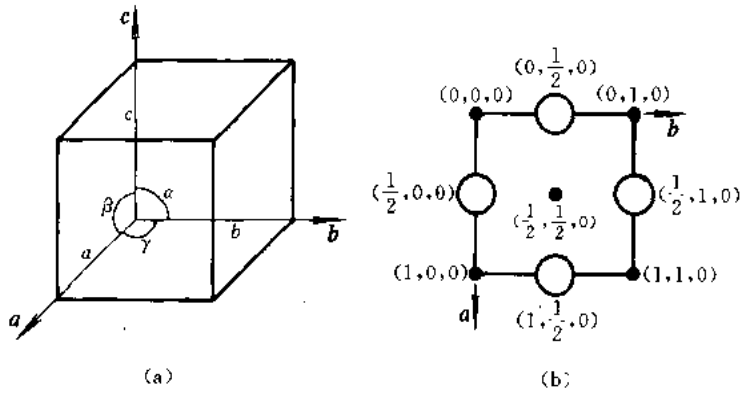


图 2.6 NaCl 晶体

(a)晶胞参数;(b)原子在二维晶胞中的坐标

与空间格子单位平行六面体一样,在一个晶体中划分晶胞的方式,可以有许多种,但是按晶体学上的要求是根据特征对称元素划分晶胞(详见 2.2.3 节)。一般选择的晶胞应该尽可能体积最小和对称性最高。体积最小就是尽可能选取最短的轴长;对称性最高就是尽可能选取立方晶胞,如不能,则尽可能按顺序选四方或六方、三方、正交、单斜晶胞等。在晶体结构中可能存在的晶胞类型共 7 种(7 个晶系),与 7 种空间格子类型相对应。表 2.1 列出了每一种晶胞类型的参数。在实际工作中三方晶胞可以按六方晶胞要求选择。

表 2.1 7 种晶胞参数

立方	$a=b=c$ $\alpha=\beta=\gamma=90°$
三方	$a=b=c$ $\alpha=\beta=\gamma\neq 90°<120°$或与六方晶胞相同
六方	$a=b\neq c$ $\alpha=\beta\neq\gamma$ $\gamma=120°$ $\alpha,\beta=90°$
四方	$a=b\neq c$ $\alpha=\beta=\gamma=90°$
正交	$a\neq b\neq c$ $\alpha=\beta=\gamma=90°$
单斜	$a\neq b\neq c$ $\alpha=\gamma=90°,\beta\neq 90°$
三斜	$a\neq b\neq c$ $\alpha\neq\beta\neq\gamma$

2.1.3 晶面指标

为讨论和研究问题的需要，晶体的空间格子可按不同方式划分成不同的平面组，各不同平面组的平面间距和平面中点阵点的密度不同(图 2.7)。同一平面组内的平面，数量无限、相互平行、间距相等、周围情况相同，每一个平面是由分布在一个平面上的点阵，按平行四边形划分的格子组成的。同理，空间格子亦可按不同方式划分成不同的直线点阵，每一组直线点阵由数量无限、相互平行、周围情况相同的直线点阵构成。

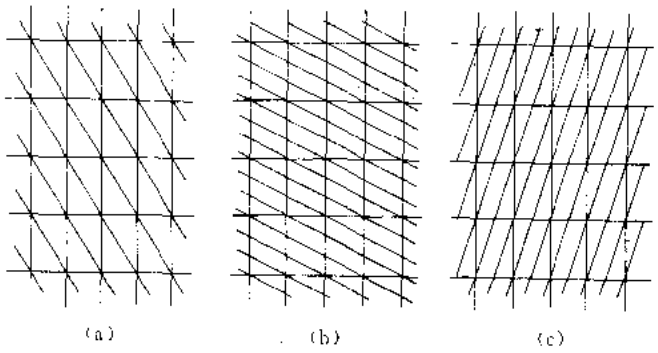


图 2.7　空间格子按 3 种不同方式划分的平面组

在晶体学的实验和理论解释中，往往需要把按不同方式划分的平面组，或者晶体外形中的某一晶面加以标记，以便区分不同的平面组，或者不同的晶面。在晶体学中任何一组晶面(晶格面)的标记都是用该组离原点最近的晶面在三个晶胞轴上的截距来表示。例如某一单位空间格子的三个轴长为 a,b,c，而某一晶面组中离原点最近的晶面在三个轴上的截距为 $a/h,b/k,c/l$，且 h,k,l 为没有公因子的互质的整数，则(hkl)为该平面组的指标，或称 Miller 指标。例如某一晶面组指标为(211)，则该晶面组离原点最近的面与 a 轴相截于$\frac{a}{2}$，与 b 轴和 c 轴分别相截在 $b/1$ 和 $c/1$，这样就确定了整个晶面组的方向和位置。

在图 2.8 中具体解释了这种指标的含意，例如(100)则表示 $h=1,k=0,l=0$ 离晶胞原点最近的面在三个晶胞轴上的截距为 $a/h=a,b/k=\infty,c/l=\infty$，即说明该面与 a 轴相截，与 b 轴和 c 轴相平衡；又如(320)晶面组是与 a 轴相截于 $a/3$，与 b 轴相截于 $b/2$，与 c 轴平行的面。其

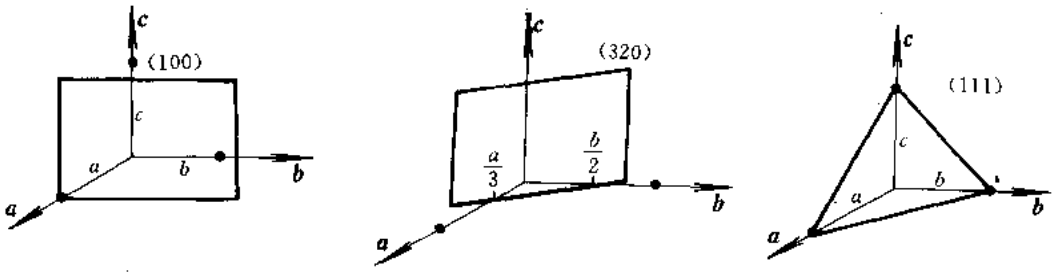


图 2.8　具有不同指标的晶面组的定向

他指标的晶面组可按同理解释。

这里必须注意，虽然 h,k,l 可以取任意整数值，但是在实际晶体生长中，只有少数格子结点密度高的晶面才能出现，(hkl)指标的数值一般不可能太高，常为 3 以下，高于 5 的值是很少的。因为指标越高，格子结点密度就越小，即分子(原子)间相互作用力越小，形成晶面的可能性也越小。[12][14]

2.2 晶体的对称性、点群、晶系、空间群

2.2.1 晶体的对称性

对称性在自然界是普遍存在的。例如，人体左右对称，花瓣常是对称地排列。所以，人们很久以来就利用对称性的概念来说明物质的一些特性。晶体结构内的分子(原子或离子)一般总是严格地按一定的对称性，排列于单位晶胞之中，使整个晶体处于能量最低、最稳定的状态。因而正确掌握对称性的概念和了解晶体结构的对称性，对于理解晶体结构和 X 射线衍射花样的特点等都是很重要的。例如 NaCl 在合适的条件下生长的晶体是立方形的，这是由于内部原子严格地按最高对称性、最密堆积排列的结果。8 个 Cl^- 对称地位于晶胞的顶角上，另 6 个 Cl^- 位于每个面的中心，而 Na^+ 对称地位于晶胞棱的 1/2 处和晶胞的中心，形成轴长相等，夹角为 90° 的立方晶胞(图 2.9)。生物大分子晶体亦不例外，例如血红蛋白分子在形成晶体时总是按一定的对称性排列于正交晶胞或单斜、四方等晶胞之中。

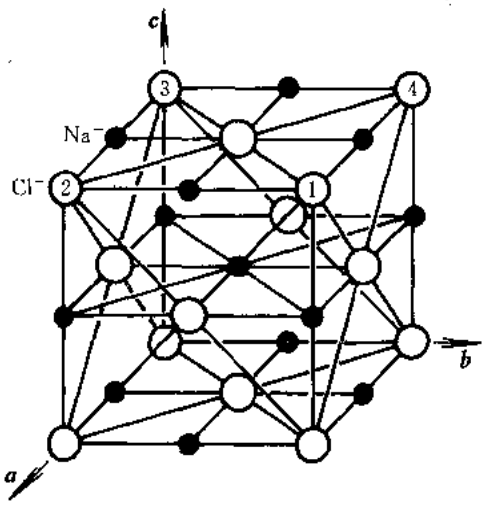


图 2.9 NaCl 晶胞
○代表 Cl^-，●代表 Na^+

由此可见，晶体的多面体外形亦是晶体内部结构对称性特点的反映，因此由晶体外形的对称性特点可推测晶体结构内部分子排列的对称性特点。这是外形与内部结构的一致性。但是，晶体外形是宏观的、有限的图形，而内部结构是分子水平的、微观的、不连续的、无限的，两者间存在着差异。因此在内部结构中存在的对称性，不能在晶体外形中全部反映出来。我们把能在外形中反映出的对称元素称宏观对称元素，不能在外形中反映出来的对称元素称微观对称元素。

对称元素是指对称动作所依据的点、线和面等几何元素。任何一个对称物体总是由两个或两个以上的等同部分组成，而通过一定的动作后，等同部分相互调换位置，整个物体恢复原状而又不改变任何内部两点间距离的这样的动作称为对称动作。

2.2.1.1 宏观对称元素

在晶体外形中可能存在的宏观对称元素有旋转轴、镜面、对称中心和旋转反轴，相应的对称动作是旋转、反映、倒反和旋转倒反。由于生物大分子是由手型分子(例如 L-氨基酸和 D-糖)构成，因此在生物分子晶体里不可能存在镜面或对称中心，因为这两种元素的动作使右或左手分子转变为左或右手分子。

2.2.1.1.1 旋转轴(n)

其对称元素是一条线，称为轴线，其对称动作是旋转。n 次旋转轴使物体绕轴旋转 $360°/n$ 均能使物体复原，n 称为轴次。最小的能使物体复原的旋转角称为基转角。在晶体中可能存在的旋转轴由于受到晶格的限制，n 只能为 1，2，3，4 和 6，相应的名称为一重轴、二重轴…六重轴。这已由数学证明。在自然界的非晶体结构中就不受此限制，例如 DNA 的碱基对是按 10 重螺旋轴排列，某些蛋白质聚合体的亚基是按 5 重旋转轴排列等等。

一个具有 n 次旋转轴的物体，围绕轴一定有 n 个相同部分，例如在 NaCl 晶体结构中(图 2.9)由(001)面中心 Cl^- 周围具有 4 个距离、环境完全相同的 Cl^- 和 Na^+。可以设想垂直于(001)面通过中心 Cl^- 处有一根轴线(对称元素)，围绕轴线顺时针方向旋轴 90°，使 Cl^-1 转到 Cl^-2，Cl^-2 转到 Cl^-3…依次类推，整个晶胞复原，即与没有旋转前完全一样，因而为四重旋转轴。以同样的道理分析 NaCl 晶胞，沿每一个对角线，都存在一个三重轴(图 2.10)。在每个晶棱中间 Na^+ 的连线位置处都存在一个二重轴等。

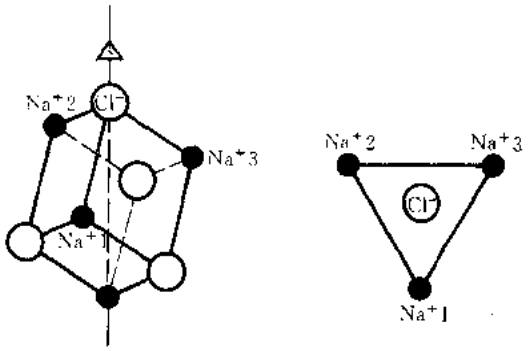


图 2.10　表示在 NaCl 晶胞中沿对角线方向上存在的三重轴对称性
△表示三重轴；右图为沿三重轴方向的投影

晶体结构中可能存在的旋转轴的轴次、符号见表 2.2。

2.2.1.1.2　镜面(m)

其对称元素是一个平面，动作是反映。物体经过镜面反映后，每一点都反映到镜面另一侧，对应点连线与镜面垂直，至镜面的距离相等，例如由图 2.11，垂直于(001)面在 1/2 轴长处可想象存在一个 m，原子 1 与原子 2 相对应，彼此垂直于 m，并等距。

2.2.1.1.3　对称中心(倒反中心)(i)

其对称元素是一个点，其对称动作为倒反。例如在[图 2.1.2(a)]NaCl 晶胞中心，可设想存在一个对称中心，对应的每一原子经过对称中心倒反以后，晶胞复原[图 2.1.2.(b)]。

2.2.1.1.4　旋转反轴($\bar{n}$)

其对称元素是轴线和轴线中心的一个点，其对称动作是绕轴旋转一个角度以后，接着依靠倒反中心进行倒反(图 2.13)。在晶体中可能存在的旋转反轴共有 5 种，与旋转轴相对应，即 5 种转轴分别与倒反中心相结合构成。但是，实际上真正独立存在的仅为四重反轴，其他几种都可以被对称中心、镜面，或者由旋转轴加镜面来代替，例如 $\bar{1}=i$，即 $\bar{1}$ 等于对称中心，$\bar{2}=m$，即 $\bar{2}$ 等于垂直于反轴的镜面。旋转反轴的轴次和符号见表 2.3。

2.2.2　点群

晶体中可能存在的宏观对称元素共有上述 4 种，而每一种晶体可能仅存在一种，或者一种

以上的对称元素。例如 NaCl 晶体外形中可能存在$\underline{4}$,$\underline{3}$,$\underline{2}$,m和i等多种对称元素，各对称元素彼此相互组合，并交于一点，组成一个对称元素系，因而称为点群(图 2.14a)。晶体中可能出现的、符合有限图形要求的组合总共有 32 种，称 32 个点群。在生物大分子晶体中由于不存在 m 和 i 对称元素，因而仅有 11 种(图 2.15)。每一个点群用一个记号表示，例如 $2/m$，表示晶体存在一个$\underline{2}$和垂直于$\underline{2}$的一个 m。

晶体结构内部的分子或原子亦是按点群的对称元素系，围绕点阵点排列的，例如 NaCl 晶体结构中原子围绕点阵点排列的对称元素系与外形中的对称元素系完全一致(图 2.14)。这样就可通过晶体外形的分析和 X 射线衍射实验来确定晶体的点群，从而获得关于分子或原子围绕点阵点排列的信息。

32 个点群的对称元素系可归纳成 7 种组合方式(表 2.4)。

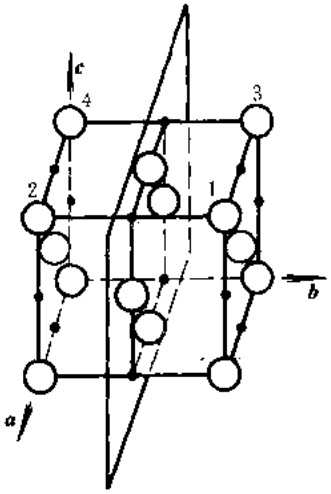


图 2.11　晶体结构中存在的对称面之一

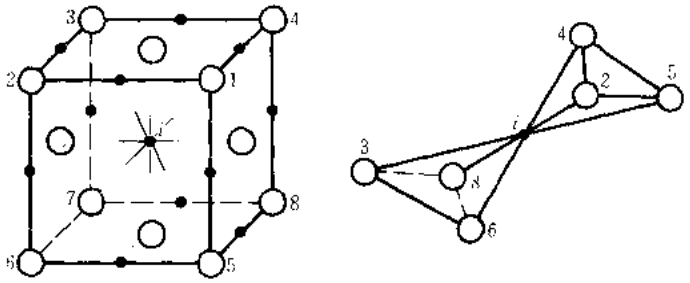


图 2.12　示意晶体结构中存在的对称中心(a)以及与对称中心联系起来的原子间的几何关系(b)

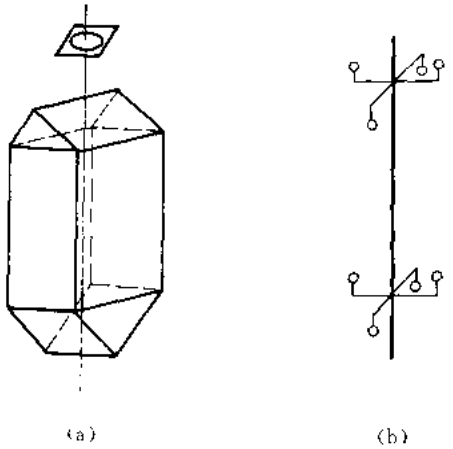


图 2.13　晶体外形中存在的四重反轴(a)以及四重反轴之图形(b)

表 2.2　在生物大分子晶体中可能存在的旋转轴和螺旋轴及其符号和记号

对称元素名称	符　号	记　号	
		平行纸面	垂直纸面
二重轴	2		
三重轴	3		
四重轴	4		
六重轴	6		
二重螺旋轴	2_1		
三重螺旋轴	3_1		
	3_2		
四重螺旋轴	4_1		
	4_2		
	4_3		
六重螺旋轴	6_1		
	6_2		
	6_3		
	6_4		
	6_5		

表 2.3　旋转反轴的轴次和符号

对称元素名称	符　号	记　号
一重反轴	$\bar{1}$	
二重反轴	$\bar{2}$	即镜面
三重反轴	$\bar{3}$	
四重反轴	$\bar{4}$	
六重反轴	$\bar{6}$	

表 2.4　符合点群组合要求的 7 种组合

n	仅有旋转轴
$\bar{n}$	仅有旋转反轴
n/m	旋转轴加上垂直于轴的镜面
nm	旋转轴加上通过轴的镜面
$\bar{n}m$	旋转反轴加上通过轴的镜面
$n2$	旋转轴加上一个垂直于轴的二重轴
n/mm	旋转轴加上一个垂直于轴的镜面和另一个垂直于镜面的镜面

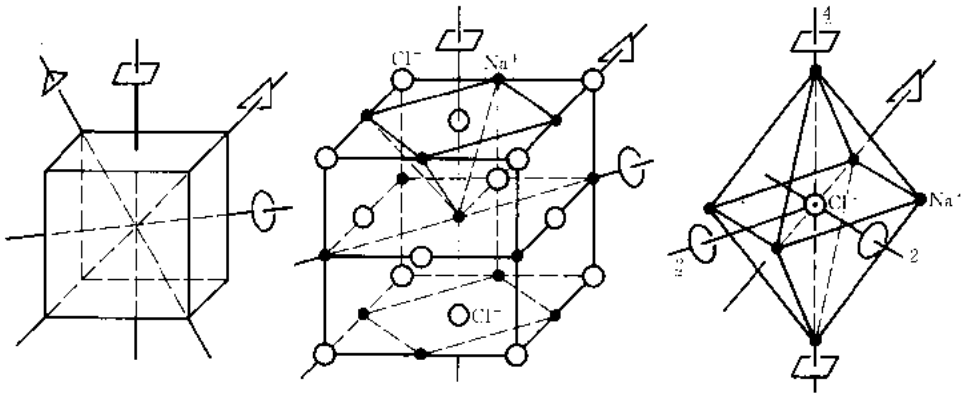


图 2.14　NaCl 晶体外形和晶胞中存在的部分对称元素

(a)晶体外形中存在的对称元素系;(b)晶胞中存在的对称元素系;(c)示意原子围绕点阵点排列的对称元素系(Cl 坐落在点阵点上)

而在生物大分子晶体结构中可能出现的组合方式仅是 n 和 $n2$,共有 11 种对称类型(图 2.15)。

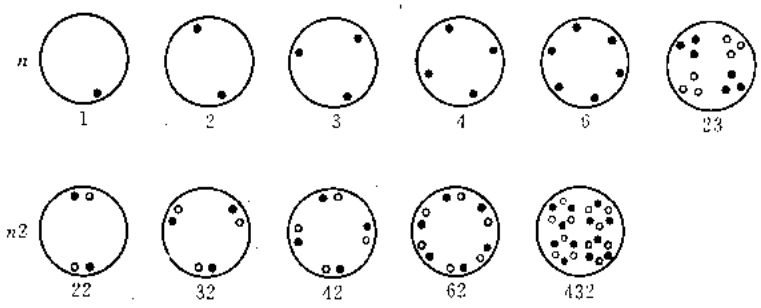


图 2.15　在蛋白质晶体中可能出现的 11 种对称类型

●表示在纸面上方;○表示在纸面下方

2.2.3　7 个晶系

根据每个点群所具有的最高轴次的旋转轴,可把 32 个点群划分成 7 个晶系,它与 7 种空间格子类型相对应。每个晶系所特有的最高轴次的旋转轴,称为特征对称元素(表 2.5)。

表 2.5　晶系的特征对称元素

晶系	格子型式	特征对称元素
三斜	P	无
单斜	P,C	一个平行于 b 的 $\underline{2}$
正交	P,C,I,F	三个互相垂直的 $\underline{2}$
四方	P,I	一个平行于 $\underline{c}$ 的 $\underline{4}$
三方	P 或 R	一个平行于 $\underline{c}$ 的 $\underline{3}$ 或一个与 a,b,c 交成等角的 $\underline{3}$
六方	P	一个平行于 $\underline{c}$ 的 $\underline{6}$ 或 $\overline{6}$
立方		4 个按立方体对角线取向的 $\underline{3}$

2.2.4 微观对称元素

2.2.4.1 平移轴

其对称元素是轴线，动作是沿某轴线方向平移一个距离。晶体结构经过平移动作后结构复原，如图 2.16 示出 NaCl 晶体结构中某一无限原子列的一部分，基元经平移 a 或 na 长度后结构复原，n 为任意正或负整数，a 为重复周期。

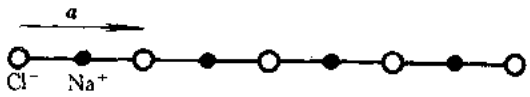


图 2.16 NaCl 晶体结构中 a 方向的平移轴

2.2.4.2 滑移面

其对称元素是一个面，动作是依靠面的反映和沿规定方向平移一定距离。滑移面由于滑移的方向和距离不同，可分为 a，b，c，n 和 d 5 种，见表 2.6。图 2.17 表示了 NaCl 晶体结构(001)面中存在 a，b 滑移面(用虚线表示)，设想在 Cl^- 与 Na^+ 之间存在一个垂直于纸面的 m，经 m 反映后，接着沿 a 方向平移 $a/2$，使 Cl^-1 移到 Cl^-2 位置，结构复原。同理在 b 方向存在一个 b 滑移面。

表 2.6 5 种滑移面的方向和平移距离

名称	a	b	c	n	d
方向	a	b	c	$a+b,b+c,c+a$	$a+b,b+c,c+a$
距离	$\frac{1}{2}a$	$\frac{1}{2}b$	$\frac{1}{2}c$	$a+b/2,b+c/2,c+a/2$	$a+b/4,\frac{b+c}{4},\frac{c+a}{4}$

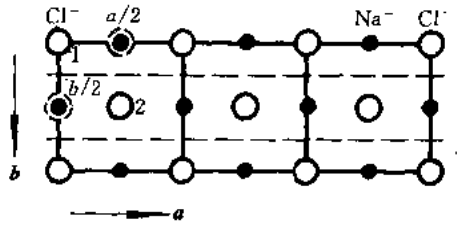


图 2.17 NaCl 晶体结构中存在的滑移面

n 滑移面(图 2.18a)，原子(0，$-b$，0)经 m 反映，接着沿 a 和 c 方向各滑移 $a/2$ 和 $b/2$，结构复原。

d 滑移面(图 2.18b)，原子经 m 反映，接着沿 a 和 c 方向滑移 $a/4$，$c/4$，结构复原。

2.2.4.3 螺旋轴(n_S)

其对称元素是一个轴，动作是首先绕轴旋转一个角度，接着沿轴方向平移一个距离，例如二重螺旋轴，首先旋转 180°，再沿该轴方向平移 1/2 单位晶轴长度(图 2.19)。在晶体结构中可能存在的螺旋轴有二重螺旋轴(2_1)，三重螺旋轴(3_S)，四重螺旋轴(4_S)，六重螺旋轴(6_S)，S 值

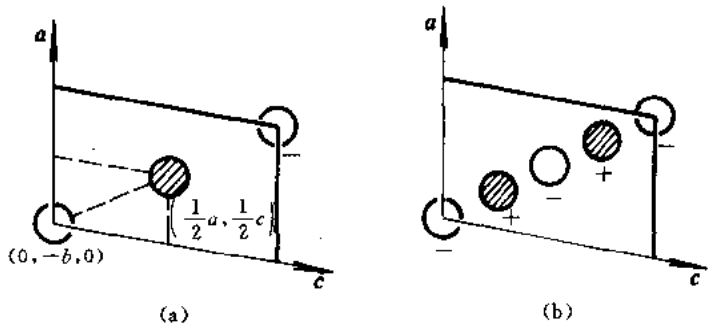


图 2.18　表示 n 和 d 滑移面

(a) n 滑移面，(b) d 滑移面。+，− 号分别表示在纸面上和下方

最大等于相应的轴次减 1，如 3_S，可以是 3_1，3_2 两种螺旋轴；4_S 可以是 4_1，4_2，4_3；6_S 可为 6_1，6_2，6_3，6_4，6_5（图 2.19）。

由于螺旋轴是旋转和平移相结合的对称元素，而旋转和平移都有一个方向问题，因此有左、右手螺旋轴之分。右手螺旋轴是把右手大姆指伸直，其余四指并拢弯曲，在此情况下，大拇指所指方向是平移方向，四指方向是旋转方向；左手螺旋轴以左手为标准。通常规定在 n_S 中，凡 $n>S<n/2$ 为右手螺旋，如 3_1，4_2，6_1，6_2；凡 $n>S>n/2$ 者为左手螺旋，如 3_2，4_3，6_4，6_5；$S=\frac{n}{2}$ 的偶次螺旋轴为中性螺旋轴。各种螺旋轴的符号见表 2.2，各种螺旋轴的图形见图 2.19。为便于比较和区别，图中同时还列出了具有旋转轴和旋转反轴的图形。

2.2.5　空间群

2.2.5.1　230 个空间群

前面讨论了晶体外形中存在的宏观对称元素和微观对称元素是相互依存的，两者又统一又有区别。首先一个晶体的宏观对称元素与微观对称元素在方向上相一致，但是，由于微观结构中存在无限多的平移轴，因而各种对称元素总是在空间作周期性的重复排列，其数无限；同时宏观的对称轴和镜面在微观结构中，还可以分别是螺旋轴和滑移面。因此属于同一点群的晶体，其内部结构的空间格子型式可以不同，对称元素亦可以不同。对称元素的不同是指在多面体外形中存在的旋转轴，在晶体结构中可能是螺旋轴，还可以是同一轴次的各种不同的螺旋轴；多面体外形中的镜面，在微观结构中还可能是各种滑移面。将确定的空间格子型式与各种可能存在的对称元素（宏观和微观）进行排列组合就可得到各种对称群，称为空间群。

例如，点群 $2/m$，属于单斜晶系，它可能出现的空间格子有 P（简单格子）或 C（001）面心格子；在结构中二重旋转轴可以是 $\underline{2}$ 或 2_1；而对称面 m，则可能为 m 或滑移面 a 和 c。经过上述空间格子的两种型式与全部 5 种对称元素组合归并可得到 6 种空间群。$P2/m$，$P2_1/c$，$P2_1/m$，$P2/m$，$C2/c$，$C2/m$。每个空间群的记号称为国际记号，大写英文字母如 P 代表点阵型式，而紧接着的是表示晶体中三个方向的对称性，如 $2/m$ 表示在 $\boldsymbol{b}$ 方向有一个 $\underline{2}$，加上垂直于此轴的 m。不同晶系在规定格子型式的大写字母后的位置上规定的晶轴方向不同。例如上例为单斜晶系第一位置代表为 $\boldsymbol{b}$ 方向，而四方晶系则为 $\boldsymbol{c}$ 方向。各晶系在三个位置上规定的晶轴方向列于表 2.7。

晶体结构中可能出现的空间格子类型共 14 种，对称元素（宏观和微观）共 7 类，如果按照

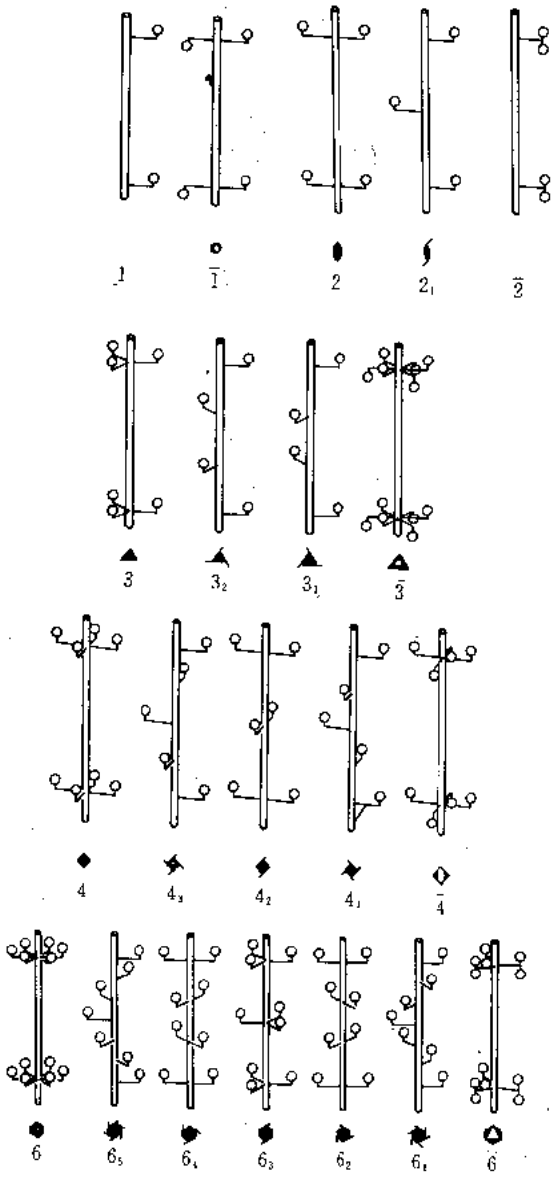


图 2.19　各种旋转轴、旋转反轴和螺旋轴之图形

格子类型与各种对称元素在一切可能的范围内进行组合，总共可得到 230 种对称元素系。与 230 种对称元素系相应的空间动作群称为 230 个空间群。

2.2.5.2　65 个非对映空间群

由于生物大分子是非对映分子，只能按非对映的空间群结晶，因而在生物大分子的晶体中可能出现的空间群，总共有 65 种，详见表 2.8。

表 2.7　国际记号中三个位置代表的方向

晶系	1	2	3
立方晶系	a	$a+b+c$	$a+b$
六方晶系	c	a	$2a+b$
四方晶系	c	a	$a+b$
三方晶系	$a+b+c$	$a-b$	—
三方晶系(按六方晶胞)	c	a	—
正交晶系	a	b	c
单斜晶系	b	—	—
三斜晶系	—	—	—

表 2.8　在生物大分子晶体结构中可能存在的空间群

晶系	点群	空间群
三斜	1	$P1$
单斜	2	$P2, P2_1, C2$
正交	222	$C222, P222, P2_12_12_1, P2_12_12, P222_1, C222_1, F222, I222, I2_12_12_1$
四方	4	$P4, P4_1, P4_2, P4_3, I4, I4_1$
	422	$P422, P42_12, P4_122, P4_12_12, P4_222, P4_22_12, P4_32_12, P4_322, I422, I4_122$
三方	3	$P3, P3_1, P3_2, R3$
	32	$P312, P321, P3_121, P3_112, P3_212, P3_221, R32$
六方	6	$P6, P6_5, P6_4, P6_3, P6_2, P6_1$
	622	$P622, P6_122, P6_222, P6_322, P6_422, P6_522$
立方	23	$P23, F23, I23, P2_13, I2_13$
	432	$P432, P4_132, P4_332, F432, F4_132, I432, I4_132$

2.2.5.3　空间群与晶体结构

每个空间群可以告诉我们，晶体结构中全部分子，或原子如何围绕点阵点按对称元素系排列，以及如何按空间格子在整个空间展开，例如空间群 $P2_1$，告诉我们分子按 2_1 轴围绕点阵点(格子结点)排列，然后按 P 格子展开。图 2.20(a)表示上述空间群在 b 方向的投影，对称元素 2_1 通过平移在 a 与 c 方向展开，(b)图表示分子围绕每一个 2_1 轴的排列，原子或分子的坐标 (x,y,z)，$\left(\bar{x},y+\frac{1}{2},\bar{z}\right)$ 是由对称元素 2_1 作用所产生的，称为等效点系。在晶体结构中凡是由

一个起始点(x,y,z)开始，通过空间群中一切对称元素作用而重复出现的一系列点的总和称为等效点系，它表示了晶胞中相同的分子，或原子分布的位置，即每一个等效点代表了在晶胞中由对称元素联系起来的、最小的重复单位，称为晶胞中一个不对称单位。每个不对称单位可由一个分子或两个以上的分子聚合体构成。

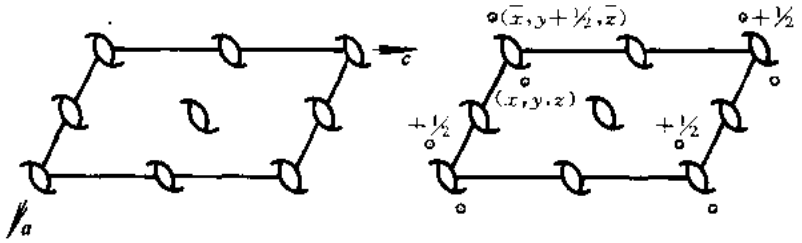


图 2.20　空间群 $P2_1$ 在 b 方向上的投影
(a)在 b 方向投影；(b)等效点系

如果通过 X 射线晶体的初步分析，确定了晶体所属的空间群，以及测定了晶体的密度，就可推算出晶胞内的分子数和每个不对称单位的组成。例如在蛋白质晶体中最小的重复单位是亚基，而每一个不对称单位可以由一个或多个亚基组成。在已知了不对称单位以后，就可按晶体所属的空间群的等效点系，大体推测分子在晶胞中的分布，例如我国科学工作者测定的猪二锌胰岛素晶体属三方晶系，晶胞参数 $a_R=11.48\text{nm}$，$\alpha_R=49.0°$，空间群为 $R3$。晶体密度为 1.245g/cm^3。根据胰岛素的分子量就可计算出每个 R 格子晶胞内含有 6 个分子，并按三重轴围绕点阵点排列，每个最小重复单位，即不对称单位包含一个二聚体胰岛素分子(图 2.21)，围绕点阵点排列的六聚体分子按 R 空间格子展开于整个晶体之中。当然整个分子的肽链走向，以至每个原子的坐标有待进一步的结构分析才能最后确定。[7][10][12]

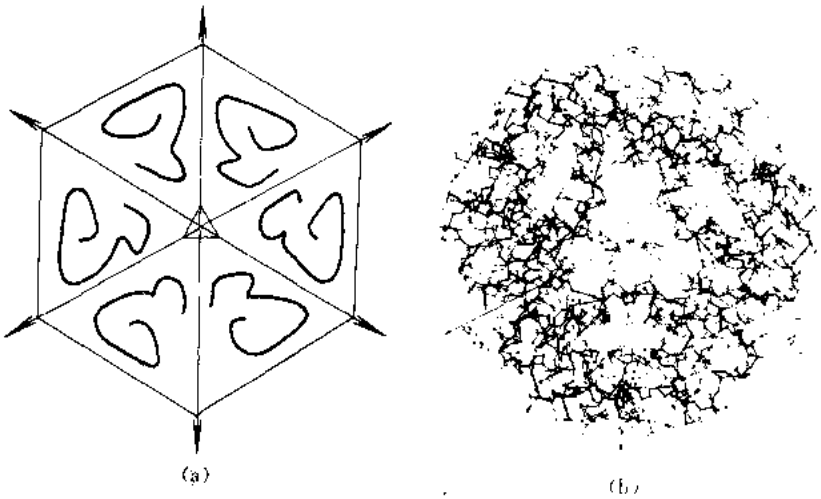


图 2.21　胰岛素晶胞沿三重轴方向的投影
(a)示意 3 个二聚体分子围绕三重轴的分布；(b)3 个二聚体分子在晶胞中的分布

第三章　晶体的X射线衍射花样

X射线晶体结构分析是利用晶体的X射线衍射现象来测定晶体及分子的结构。而X射线衍射可以简单地理解为当一束平行的X射线投射到晶体上时，大部分入射线穿过晶体沿原方向前进，而部分射线却偏离了入射方向（图3.1）。如果设法用X光胶片把这种衍射现象记录下来，就可得到晶体的衍射花样。所谓衍射花样是指衍射线在空间的几何分布和衍射线的强度。因而每一张记录衍射花样的胶片包括了衍射点和强度的分布。图3.2示意了本实验室记录的斑头雁氧合血红蛋白晶体的部分衍射花样，在图中衍射点

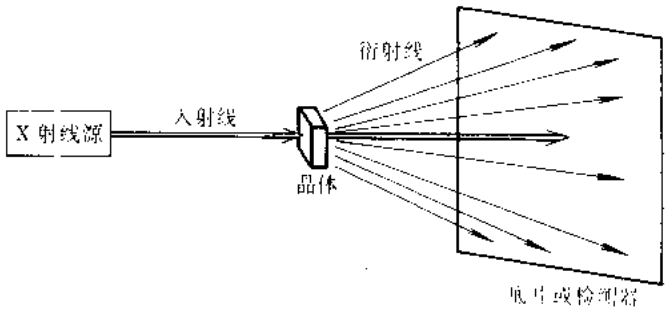


图3.1　晶体的衍射

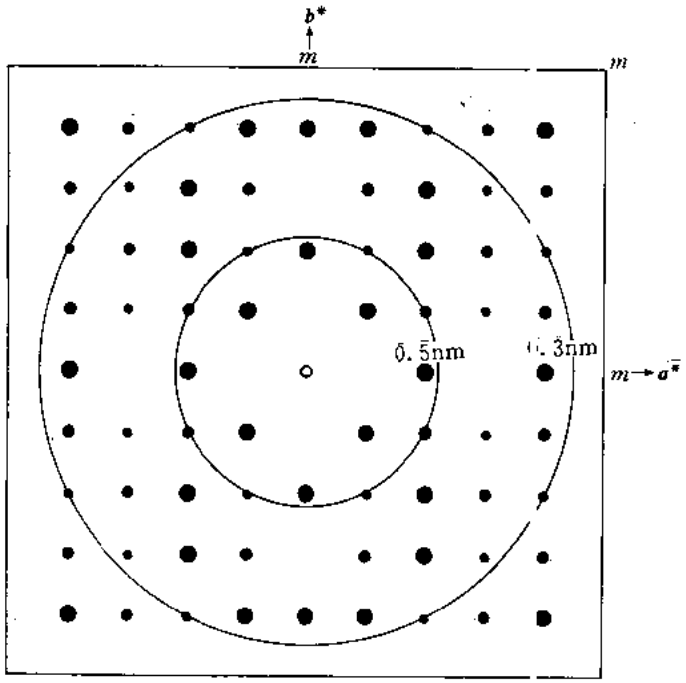


图3.2　示意衍射花样上衍射点及其强度分布
圆圈上的数值表示分辨率范围

在a^*和b^*方向上按一定间距并对称地分布，全部衍射点围绕着中心呈四重轴对称性排列，而衍射点的强度有弱、有强，有的系统地消失。在这种花样中，点的分布特征是由斑头雁血红蛋白分子在晶体结构内重复排列的方式所决定的，而点的强度分布特征是由分子结构本身的特征所决定的，即衍射花样的特征是由晶体结构所决定。

本章将着重讨论衍射花样为什么是由分立的一个个点构成，以及点的分布与晶胞大小、形状的关系，点的强度与晶胞中原子分布的关系。

3.1　衍射线的分布

3.1.1　一个原子的散射

晶体是由各种原子组成的，而原子是由电子和原子核构成的，当X射线波投射到原子上时，每个电子在外加电磁场的作用下，被迫发生振动。振动着的电子向四面八方发射次生电磁波，成为新的电磁波源（图3.3）。次

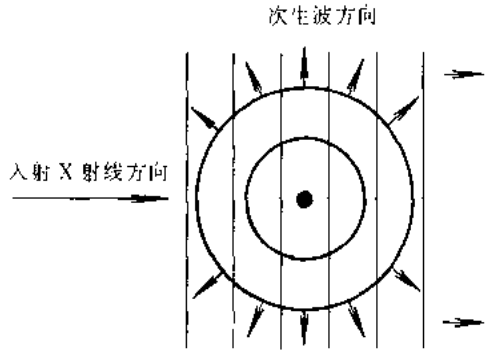


图 3.3　示意入射波与散射波之间的方向关系

生波的频率、波长与入射线的频率、波长相等。由于原子核的质量比电子的质量大得多(e/m～1∶1840),所以在讨论原子的散射时,可以忽略核的振动而仅仅考虑电子的振动。另外,为讨论衍射问题的方便,在此近似地假定原子核外的电子都集中在一个点上,忽略各电子散射波之间的相互干涉。这样,把整个原子看成产生次生电磁波的一个波源,它的散射能力决定于原子的电子数的多少。

3.1.2　无限原子列的衍射方向

假定一个由同种原子组成的无限原子列,原子的重复周期为 a,即相当于一列直线点阵中,每个点阵点上仅有一个相同的原子。当 X 射线冲击到该原子列上时,每个原子各自发射次生球面波,由于次生波具有相同的频率和波长,相互必然发生干涉,而干涉的结果必定在某些方向上波的强度加强,而在某些方向上波的强度减弱。为讨论的方便,首先研究在该原子列上

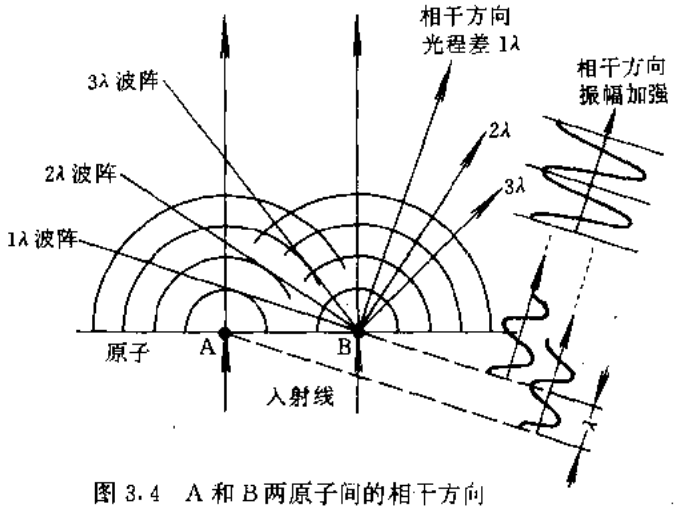


图 3.4　A 和 B 两原子间的相干方向

的任意相邻两个原子间的干涉方向。假如任意相邻两原子 A 和 B(图 3.4),它们各自向四周发射次生波并相互发生干涉,干涉的结果产生合成波,而合成波的强度在不同方位上是不同的,仅在两个次生波的光程差为波长的整数倍的方向,强度最大,即在光程差为 $1\lambda, 2\lambda, 3\lambda, \cdots, n\lambda$ 等的方向上强度最大,而在其他方向上光程差一定不是波长的整数倍,相干后的合成波的强度

必然小于整数倍的方向，而在光程差为 $1/2\lambda$ 的方向上，相干后强度为零(图 3.4)。

同理在无限原子列的情况下，当各原子发射的次生波相互间干涉的结果，仅在相邻两原子间光程差为波长的整数倍方向上，强度最大，而在其他方向(非整数倍方向)上由于各原子间次生波相互干涉而使强度为零。因此，只有 $\Delta=n\lambda$ 的方向是无限原子列的衍射方向，Δ 为光程差，n 为任意整数。这里再强调一下，所谓衍射方向就是指由重复周期(此处为 a)联系起来的相邻两原子，或基元之间次生波相干后合成波强度加强的方向，即 $\Delta=n\lambda$ 的方向。而其他方向，即 $\Delta\neq n\lambda$ 的方向由于在无限原子列中各原子间的相干而使衍射线的强度为零。

3.1.3 无限原子列的衍射线的分布

假设一束平行的 X 射线，与上述无限原子列成 90°(α°)的夹角入射到该原子列上，各原子散射次生波，并相互干涉，假定在与原子列成 ϕ 角的方向上产生衍射。根据产生衍射的条件，相邻两次生波的光程差应等于波长的整数倍。由图 3.5，两次生波之间光程差为：

$$\Delta = BC$$

而

$$BC = a\cos\phi$$

则

$$\Delta = a\cos\phi$$

当 $\Delta=h\lambda$ 时，即 $a\cos\phi=h\lambda$ 时，衍射产生，此处 a 为原子的重复周期，h 为任意整数，称为衍射指标，是原子列上任意相邻两原子散射次生波之间的光程差，以 λ 为单位，λ 为 X 射线的波长。ϕ 称为衍射角。

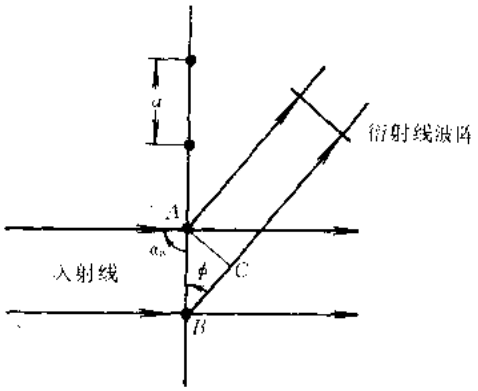


图 3.5 无限原子列中相邻两原子间的相干方向

上述等式称之为劳厄(Laue)条件，由任何无限原子列所产生的一切衍射，都必须满足方程。

由公式

$$a\cos\phi = h\lambda$$

$$\cos\phi = h\lambda/a$$

当 h 和 λ 为某一定值时，ϕ 亦为常数。ϕ 角规定了衍射线的方向，凡是与该原子列成 ϕ 角的方向上，均符合劳厄条件，产生衍射线。而 ϕ 角的大小决定于周期 a 的大小。

当 $h=0$ 时，称为 0 级衍射，这时 $\cos\phi=0$，$\phi=90°$，即所有 0 级衍射线都分布在与原子列成 90°的圆锥面上，而圆锥的顶角大小与周期 a 的大小无关。当 $h=\pm1$ 时称为一级衍射，这时 $\cos\phi=\pm\lambda/a$；$h=\pm2$ 时，称为二级衍射，$\cos\phi=\pm2\lambda/a$；依次可得到不同 ϕ 值。以 ϕ 值作图可得到各级衍射线的分布(图 3.6)。由图可见所有衍射圆锥都分别位于入射线两侧的同轴圆锥上，第二级的圆锥顶角小于第一级圆锥的顶角，第三级小于第二级，以此类推。

如果 X 射线以与原子列成任意夹角(α_0)的方向入射到原子列上，则由图 3.7 可推得光程差：

$$\Delta = \overline{BC} - \overline{AD}$$

$$\overline{AD} = a\cos\alpha_0$$

$$\overline{BC} = a\cos\phi$$

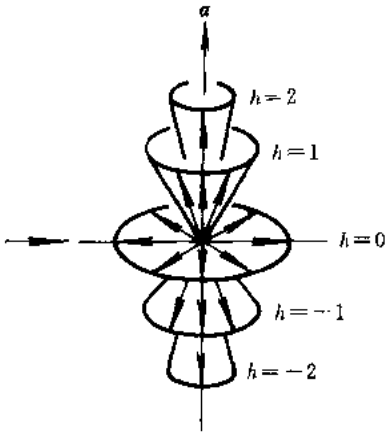


图 3.6　无限原子列的各级衍射线的分布

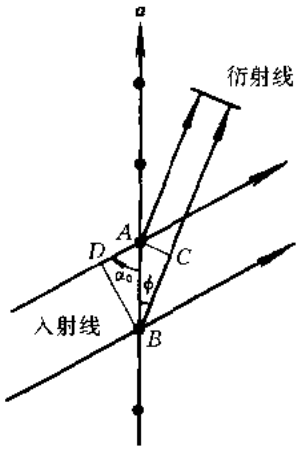


图 3.7　入射线与原子列成任意角时无限原子列的衍射方向

$$\Delta = a\cos\phi - a\cos\alpha_0 = h\lambda$$

$$\cos\phi = \cos\alpha_0 + h\lambda/a$$

同样，ϕ 角规定了衍射线的方向，当 h，λ 以及入射角 α_0 一定时，ϕ 角大小决定于 a 的大小。

衍射圆锥同样分布在入射线的两侧，二级圆锥的顶角小于一级，三级小于二级…等等。但是由于当 $h=\pm1$ 时，$\cos\phi=\cos\alpha_0\pm\frac{\lambda}{a}$ 的值不同，所以两圆锥顶角 ϕ 的大小不同，$h=1$ 时的 ϕ 角小于 $h=-1$ 时的 ϕ 角(图 3.8)。

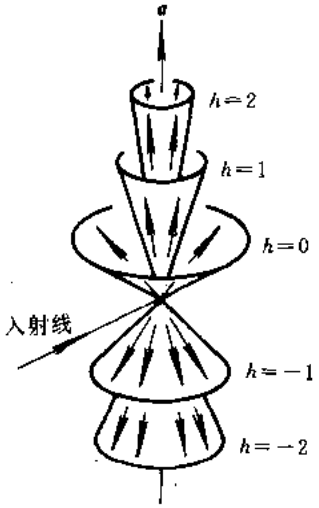


图 3.8　入射线与原子列成任意角时各级衍射线的分布

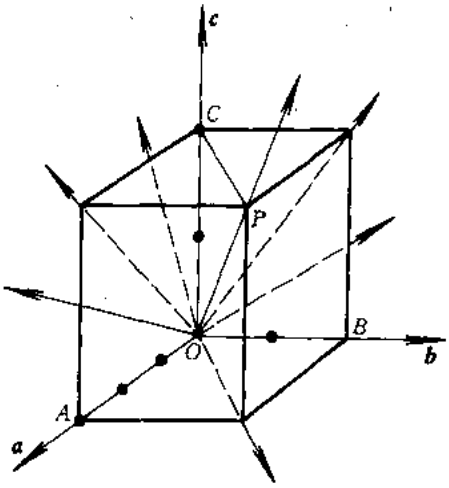


图 3.9　通过原点 O 的直线点阵以及 $\overrightarrow{OP}$ 向量分解为 $\overrightarrow{OA}$，$\overrightarrow{OB}$ 和 $\overrightarrow{OC}$ 的情况

当 $h=0$ 时，$\cos\phi=\cos\alpha_0$，形成零级衍射圆锥，ϕ 角小于 180°，它的大小决定于入射角(α_0)的大小，而与原子列周期大小无关。

综上所述，不论 X 射线以多大角度入射到无限原子列上，衍射角(ϕ)的大小决定了衍射线的方向，当 λ, h, α_0 唯一确定时，ϕ 角的大小决定于原子列周期(a)的大小，a 大则 ϕ 角小，a 小则 ϕ 角大，二者成反比关系；但是零级衍射圆锥角的大小与 a 无关，仅决定于入射角的大小。各级衍射角之间的变化是不连续的、分立的变化。在某一级衍射中如 2 级，则由 a 联系起来的全部相邻原子间的光程差都是 2λ，因而散射线的步伐完全一致。

3.1.4 晶体的衍射方向

任何一个三维晶体格子总可以由三个相交，而不共面的直线点阵决定，相应地晶体结构同样可以由三个相交而不共面的原子列决定。这样，整个晶体结构对 X 射线所产生的衍射条件也将由上述的三个原子列的衍射条件共同决定。假如以 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 表示三个轴的方向，并相交于 O 点，三个轴之间的夹角为 α, β, γ，三个方向原子列的重复周期分别为 a, b, c，则由上述 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 三轴所规定的具有一定晶胞参数的晶体结构，可以看成由通过共同原点 O 的无数直线点阵式的原子列组成(图 3.9)。当 X 射线以任何入射角冲击到该晶体时，从理论上讲，晶体结构内任何一原子列如果满足衍射条件(衍射方向决定于该原子列的重复周期)都可产生自身的衍射，但是由于晶体结构是由上述的无数直线点阵式的原子列构成的，因此作为整个晶体要在某一方向上产生衍射，必须使晶体的全部原子列同时满足该方向上的劳厄条件。但是由于晶体结构内任何一原子列上，从原点 O 至任一原子的向量，都可分解为 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 三个轴方向上的向量(图 3.9)。因此整个晶体的衍射将由三个轴的衍射条件来决定。而每个轴的劳厄条件可按 3.1.3 节所讨论的方式推出，在形式上完全类似，具体如下：

$$a(\cos\alpha - \cos\alpha_0) = h\lambda$$
$$b(\cos\beta - \cos\beta_0) = k\lambda$$
$$c(\cos\gamma - \cos\gamma_0) = l\lambda$$

这一联立方程式称之为劳厄方程式，它规定了任何一个三维空间的晶体结构的衍射条件。式中 $\alpha_0, \beta_0, \gamma_0$ 分别表示入射线与三个原子列的交角；而 α, β, γ 分别为衍射线与三个原子列之间的夹角(图 3.10)。h, k, l 均为整数，称为衍射指标。某一具体的 hkl 值规定了某一特定的衍

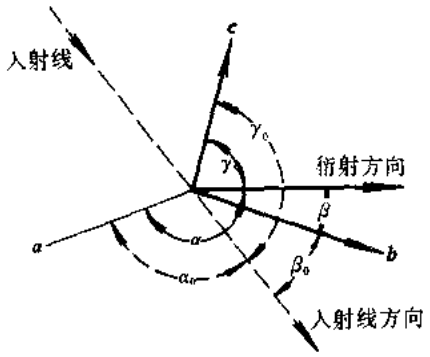


图 3.10 入射线、衍射线与晶胞参数间的关系

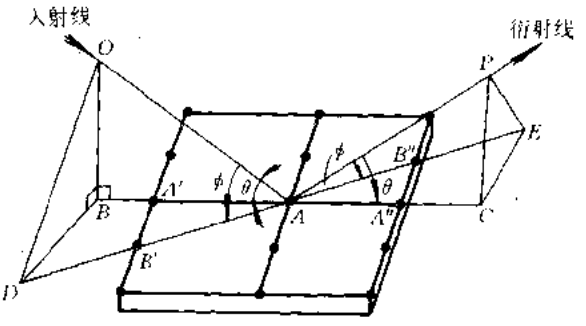


图 3.11 平面点阵中各直线点阵产生的共同"反射"线

射线的方向，不同的 hkl 值规定了衍射线的不同方向，例如 231 衍射，即表示在 $\boldsymbol{a}$ 方向由重复周期 a 联系起的任意相邻两原子的光程差为 2λ，同理在 $\boldsymbol{b}$ 和 $\boldsymbol{c}$ 方向由 b 和 c 联系起来的相邻

两原子间光程差为 3λ 和 1λ。衍射的不同级别，说明了相应原子列上由重复周期所联系起来的任意相邻两原子间散射的次级波间的光程差(以 λ 为单位)不同。

由于衍射指标 hkl 都是正或负的整数，因此衍射线的分布必然是一些分立的线，记录下来的是一些分立的点。

一个晶体要产生衍射，必须同时满足三个方向上的劳厄条件，因为各衍射角(α,β,γ)之间有一定的关系。当 λ 和衍射指标 hkl 为一定值时，衍射线的方向除决定于晶胞轴长 a,b,c 以外，还决定于晶胞轴之间的夹角。不同的夹角与入射线和衍射线之间的交角不同，即在公式中的 $\alpha_0,\beta_0,\gamma_0$ 和 α,β,γ 值不同，这说明晶体的衍射线的方向和在空间的分布决定于晶胞的形状和大小。由此可以设想分析衍射花样上点的分布，可推测晶胞的形状和大小。

3.1.5 布拉格(Bragg)方程

任何一个三维的晶体结构，总可以划分成大量的、不同取向、不同间距的相互平行的平面格子式的结构。1912 年布拉格(W. L. Bragg)从某一组平面格子式结构的衍射特性来讨论整个晶体的衍射，并建立了衍射方程，称布拉格方程。

在 3.1.3 节中讨论无限原子列的衍射方向时，已经指出了零级衍射线的衍射角总是等于入射线的入射角，衍射角的大小决定于入射角的大小，而且与原子的重复周期无关。借助光学中对反射的定义，入射角等于反射角，可以把无限原子列的零级衍射看成是“反射”。

一个平面格子式的结构总可以划分成各种不同方式的直线点阵。当 X 射线以某一角度入射到该平面上时，平面内的每个直线原子列将产生各自的零级衍射圆锥，但是一般情况下它们的衍射方向是不同步的，要得到共同的零级衍射方向，必须使所有衍射圆锥相切于一个方向。下面将证明这样的共同零级衍射线是整个格子平面的“反射”线。假如入射线 OA 以垂直于平面的上方(即 OA 与平面法线在同一平面内)打到周期为 a 的原子列 $A'A''$上，入射角为 θ，零级衍射角亦为 θ(图 3.11)，衍射线为 AP，OAB 平面垂直于平面点阵，即$\angle OBA=90°$。平面格子中任意一直线点阵 $B'B''$，与 OA 入射线构成的入射角为 ϕ，零级衍射角亦为 ϕ，零级衍射线亦为 AP。AP 衍射线为直线点阵 $A'A''$与 $B'B''$的零级衍射圆锥的相切线，并且由锥体$\triangle OBAD$ 与 $\triangle PCAE$ 完全相似而证明 AP 线在 OBA 同一平面内。同理可以证明平面内任何方式的、不同的直线点阵的零级衍射圆锥都相切于 AP 线。由此证明所有直线点阵所产生的共同零级衍射线 AP 仿佛是整个平面格子对入射线的反射，即入射角 θ 等于衍射角，入射线与衍射线构成的平面垂直于平面格子(反射面)。整个平面内不同周期联系的任何相邻两原子的入射线和衍射线的光程相等(图 3.12)，而整个晶体的衍射决定于相邻两平面的“反射”线的光程差，当光程差等于波长的整数倍时，衍射产生。

假设晶体的任意一组平面中的 AB 面(图 3.13)的入射线和所产生的“反射”线较之由相邻平面 $A'B'$的入射和“反射”线要多走 $\overline{EO}+\overline{OF}$ 一段距离，即两者之光程差 Δ 为：

$$\Delta = \overline{EO} + \overline{OF} = \overline{OD}\sin\theta + \overline{OD}\sin\theta = 2\,\overline{OD}\sin\theta = 2d\sin\theta$$

式中 d 为平面间距，θ 为入射角。要使衍射产生，必须满足相邻平面间的光程差为波长的整数倍，即

$$2d\sin\theta = n\lambda$$

式中 n 为自然数，称为反射级数。

这一方程式称为布拉格公式，它是晶体学中最基本的公式之一。公式中的 θ 角特别称之为

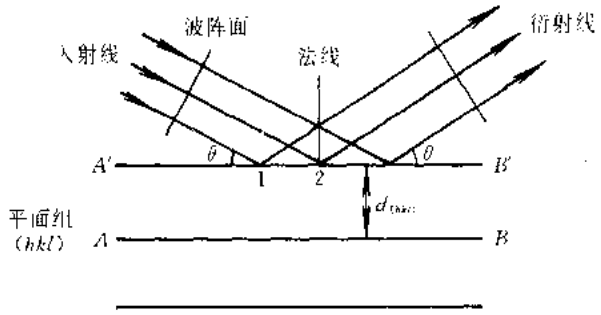


图 3.12　平面内任一周期相联系的相邻两原子 1,2 的光程相等

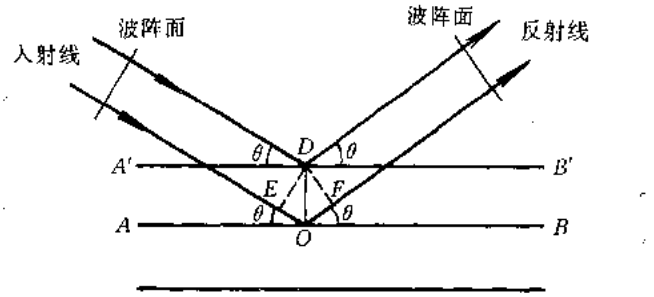


图 3.13　平面组对 X 射线的"反射"

布拉格角，或称为衍射角。这个公式确定了任何衍射线在空间的位置，并由公式 $2d\sin\theta = n\lambda$ 得 $\sin\theta = \frac{n\lambda}{d}$，$\theta$ 角的大小决定了"反射"线的方向，当 n，λ 一定时，θ 角的大小决定于 d 的大小，并互成反比，即平面组(hkl)的间距 $d_{(hkl)}$大，则 θ 角小。对某一特定的平面点阵组(hkl)来说，n 值不同，如 $n=1,2,3\cdots$，则相应的衍射角分别为 $\theta_1,\theta_2,\theta_3\cdots$，即衍射方向也不同。

对某一平面组来说，带有公因子的平面指标$(nhnknl)$是与(hkl)平面组平行的一组面，当相邻两平面点阵间的光程差为 $n\lambda$ 时，那末沿着 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 三个轴，由单位平移向量$(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$联系起来的，或沿着任何平行于轴方向的相邻两点阵点间的光程差为 $nh\lambda,nk\lambda,nl\lambda$，即平面指标$(hkl)$与劳厄方程中的衍射指标 hkl(不带括号)间的关系为$(nh)=h,(nk)=k,(nl)=l$，注意(nl)中的 l 是晶面指标，而等式右边的 l 是衍射指标。由图3.14，当(111)平面组相邻两平面间的光程差 $n=1$时，则(0,0,0)点阵点与(0,B,0)点阵点间的光程差为 $k\lambda=1\lambda$，而对(222)平面组来说，当 $n=1$时(0,0,0)与(0,B,0)点阵点间的光程差 $k\lambda=2\lambda$，同理可理解在 $\boldsymbol{a}$ 与 $\boldsymbol{c}$ 方向上相邻两点阵点间的光程差与相邻两平面间的光程差 $n\lambda$ 之间的关系。[2][10][12][13]

3.1.6　衍射线的分布与倒易晶格

劳厄方程和布拉格公式都可预示衍射方向，并由布拉格公式$2d\sin\theta=n\lambda$ 可以算出决定衍射方向的 θ 值，而 θ 的大小与晶体间距 d 成反比，即若 d 大则 θ 值小，因而衍射花样中衍射点分布密集；而若 d 小则点分布松散。由此，如能够建立起 θ 与 d 之间的定量关系，那末就有可能从已知晶体的不同 d 值，推知 θ 的不同值，从而预示晶体的衍射线的几何分布。这对于解释晶体的衍射现象和理论计算等，无疑是非常有益的。

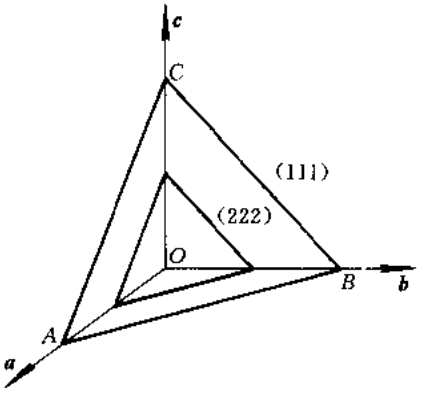


图3.14　平面指标(hkl)、衍射指示 hkl 与 n 间的关系
$OA=a, OB=b, OC=c$

为便于讨论，我们以向量的形式来表示布拉格公式，为此用单位向量 s_0 和 s 分别代表入射线和衍射线的方向，并令 $|s|=|s_0|=\frac{1}{\lambda}$，$\lambda$ 为 X 射线的波长。根据公式 $2d\sin\theta = n\lambda$，当晶体产生衍射时，入射线与"反射"面间的夹角为 θ，则入射线与"反射"线之间的夹角为 2θ，入射线与"反射"线在同一平面，并垂直于反射面。

由图3.15可看到：$\boldsymbol{S}=\boldsymbol{s}-\boldsymbol{s}_0$，并垂直于反射面 OP，$|S|=CP+PD$，由 ΔOPD 得 $\sin\theta=PD/|s|$，$PD=\sin\theta|s|=\sin\theta\frac{1}{\lambda}$，由 ΔCOP 得 $CP=\sin\theta\ |s_0|=\sin\theta\frac{1}{\lambda}$，则 $|S|=2\sin\theta\frac{1}{\lambda}$。

由公式 $2d\sin\theta=n\lambda$，当 $n=1$ 时，则 $\frac{1}{d}=\frac{2\sin\theta}{\lambda}$。比较两式得：$|S|=\frac{1}{d}$。当 n 为任意整数时，$|S|=n/d$。

$\boldsymbol{S}$ 向量的方向始终垂直于反射面，而当 n 为某一定值时，它的大小反比于"反射"面间距 d。故称 $\boldsymbol{S}$ 向量为倒易向量。如果在 $\boldsymbol{S}$ 向量的末端 D 处用一个点表示，则该点称为倒易点，它与 O 的连线方向，即为衍射线的方向。当 n 为不同值时，$\boldsymbol{S}$ 向量的方向不变，仅大小改变，则规定了不同级的衍射线的方向。

通过倒易向量，就可预示任何一个晶格的倒易晶格，从而知道衍射线(衍射点)的分布特点：

(1) 一组晶格面的倒易点的分布是一列点

设任意一组晶格面的间距为 $d_{(hkl)}$，则根据 $\boldsymbol{S}=n/d$，n 为不同值时，得相应的一组向量，该组向量都垂直于晶格面，其大小为 $1/d, 2/d, \cdots, n/d$。在每个向量的末端用一个点来代表，就得到一列点(图3.16)。这列点称为倒易点列。

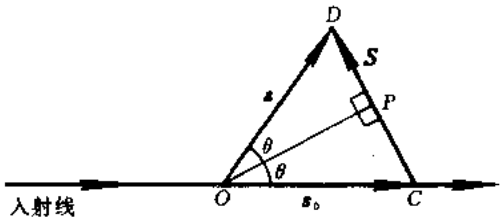


图3.15　布拉格公式的向量表示

(2) 平面晶格的倒易点的分布是一个平面格子

例如，设一个单斜晶格，a 与 c 轴斜交，夹角为 β，而 b 轴垂直于 ac 平面。图3.17表示该晶格沿 b 方向的投影，a 与 c 在纸平面内，β 角等于 $\angle aOc$。单位格子沿 b 轴的投影就是平行四边形 $OABC$，$\overrightarrow{OA}$ 等于晶胞的 a 边，而 $\overrightarrow{OC}$ 边为 c 边。将该单位格子向 a 与 c 方向重复平移就构成平面格子。在图上格子结点用小的实心圆代表。通过格子点可以划分各种平面组，在任何一平面组内，其相邻面的间距是一定的。按照 $|S|=\frac{1}{d_{(hkl)}}$，就可求出各组面的倒易点。例如相当于平面组(100)的倒易点在图中是垂直于(100)面的 $\overrightarrow{OP}$ 线的顶点，$|OP|=\frac{1}{d_{(100)}}$，用方块表示，倒易点指标为100。用同样的方法可求出其他平面组的相应的倒易点，平面组(001)的倒易点为 Q 点，平面组(101)的倒易点为 B' 点，$OPB'Q$ 组成一个单位平行四边形，称为倒易单位平面格子，沿

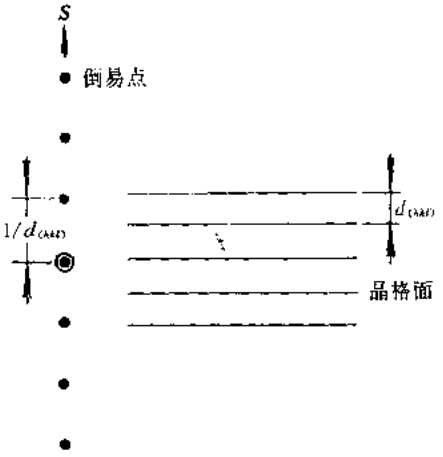


图3.16　一组晶格面的倒易点列

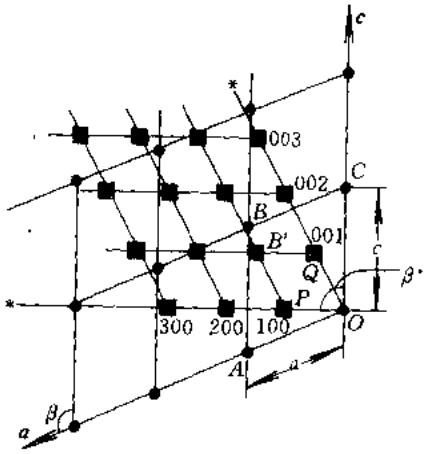


图3.17　平面晶格与倒易晶格之间的关系

$\overrightarrow{OP}$,$\overrightarrow{OQ}$方向重复平移该单位格子就得到整个倒易平面格子。由此证明一个平面格子的倒易点的分布仍然是一个平面格子。

(3) 三维晶格的倒易点的分布是三维格子

同理,在上述单斜平面晶格的基础上可推出平面组(010)的倒易格子点,从而证明单斜晶格的倒易晶格仍然是单斜的三维格子。如果倒易单位格子的参数用 a^*,b^*,c^*,α^*,β^*,γ^* 表示,则 c^* 垂直于晶格的 ab 平面,a^* 垂直于 bc 平面,b^* 垂直于 ac 面,且 $a^*=1/d_{(100)}$,$b^*=1/d_{(010)}$,$c^*=1/d_{(001)}$(图3.18)。将倒易单位格子向各方向重复就构成单斜晶格的倒易格子。

由上面的讨论可得出如下结论:任何一组平面的倒易点的分布是一列点;任何二维晶格的倒易点的分布仍是二维格子;任何三维晶格的倒易点的分布仍是三维格子。两种晶格间的大小关系是倒易关系,由此可用数学公式来定义晶格与倒易晶格向量之间的关系:

表 3.1　单位晶胞与倒易晶胞间的关系

正交晶系

$a^*=\frac{1}{a}$,$b^*=\frac{1}{b}$,$c^*=\frac{1}{c}$

$\alpha^*=\beta^*=\gamma^*=90°$

$V^*=\frac{1}{V}=a^*b^*c^*$

单斜晶系

$a^*=\frac{1}{a\sin\beta}$

$b^*=\frac{1}{b}$

$c^*=\frac{1}{c\sin\beta}$

$V^*=\frac{1}{V}=a^*b^*c^*\sin\beta^*$

三斜晶系

$a^*=\frac{bc\sin\alpha}{V}$

$b^*=\frac{ac\sin\beta}{V}$

$c^*=\frac{ab\sin\gamma}{V}$

$V^*=\frac{1}{V}=a^*b^*c^*(1-\cos^2\alpha^*-\cos^2\beta^*-\cos^2\alpha\gamma^*+2\cos\alpha\cos\beta\cos\gamma)^{1/2}$

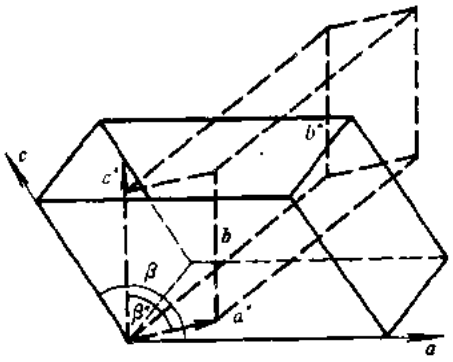


图3.18 单斜晶胞与倒易晶胞间的关系

图中虚线表示倒易晶胞

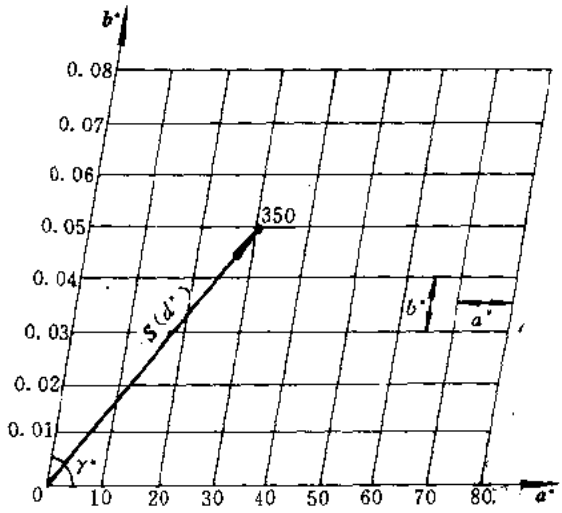


图3.19 倒易格子结点(点阵点)至原点的距离$|S|$

$$a\cdot a^*=1\quad a\cdot b^*=0\quad a\cdot c^*=0$$
$$b\cdot a^*=0\quad b\cdot b^*=1\quad b\cdot c^*=0$$
$$c\cdot a^*=0\quad c\cdot b^*=0\quad c\cdot c^*=1$$

由定义式可知,$a^*=1/d_{(100)}$,$a^*\perp bc$ 面;$b^*=1/d_{(010)}$,$b^*\perp ac$ 面;$c^*=1/d_{(001)}$,$c^*\perp ac$ 面。在直角坐标系中,$a^*=1/a$,$b^*=1/b$,$c^*=1/c$。表3.1列出了单斜、正交等晶系的晶格与倒易晶格之间的关系。其他晶系的晶格与倒易晶格间的关系可从《国际晶体学表》中查到。

由晶胞与倒易晶胞之间的关系,可以从已知的晶胞参数来预示倒易点的分布(衍射点的分布),并对每个倒易点进行指标化,反之亦可由已知衍射点的分布来推算晶胞的大小和形状。

假如以$|S|$来表示从倒易格子原点至任何一倒易点的距离,并以 hkl 为该点的指标(图3.19),则

$$|S|=(h^2a^{*2}+k^2b^{*2}+l^2c^{*2}+2hka^*b^*\cos\gamma^*+2hla^*c^*\cos\beta^*+2klb^*c^*\cos\alpha^*)^{1/2}$$

而
$$|S|=\frac{1}{d_{(hkl)}}=\frac{2\sin\theta}{\lambda}=d^*$$

由上述关系可以计算从任一倒易点至原点的距离。

3.1.7 晶体的衍射方向与反射球

晶体衍射点的分布可以用倒易晶格来描述,但是,某一衍射能否产生,还得看是否满足劳厄方程或布拉格公式。1921年爱华德(Ewald)提出用反射球的概念来描述衍射产生的条件。

在图3.20中,EO 为入射线,以 C 为圆心,$CO(CO=1/\lambda)$为半径作一个圆。假设在 C 处有一组晶面,以 CG 代表,按布拉格公式,则$\overrightarrow{CP}$为"反射"线方向,$\overrightarrow{CO}$为入射线方向,$\overrightarrow{OP}$即为 S 向量,

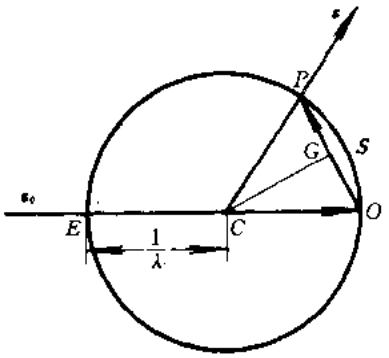


图3.20　以$\frac{1}{\lambda}$为半径的反射球

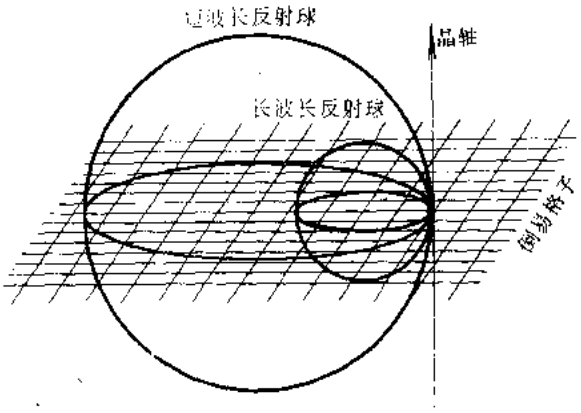


图3.21　两种不同波长的反射球

点 P 为倒易晶格点。按布拉格公式的向量表示，P 点必定满足布拉格公式而产生衍射，方向为 $\overrightarrow{CP}$。同理可以证明在球壳上任何一点都可满足布拉格公式而产生衍射，故把$1/\lambda$为半径的球称为反射球。反之如果要使某一衍射产生，就必须设法转动晶体，使倒易晶格点与反射球相碰而满足布拉格公式。

反射球的大小直接取决于 X 射线的波长，波长短则反射球大，波长长则反射球小(图3.21)。在$2/\lambda$的直径范围之内的所有倒易点都可能与反射球表面相碰而产生衍射，所以在已知波长的条件下，可以计算可能收集的最多衍射点数 N，如果以简单晶胞为例，则

$$N = \frac{(4/3)\pi(2/\lambda)^3}{\text{倒易晶胞体积}}$$

3.2　衍射线的强度

3.2.1　一个原子的散射

根据电磁理论，电子散射波在与入射线夹角为2θ处的强度

$$I_{2\theta} = I_0 \frac{ne^4}{2r^2m^2c^4}(1 + \cos^2 2\theta)$$

式中 I_0为入射线的强度，n 为电子数，r 为离散射中心的距离，e 为电荷，m 是电子质量，c 为光速，$(1+\cos^2 2\theta)$项是 X 射线部分偏振的修正。由公式可见，散射强度与质量成反比，而与电子数成正比。因此在考虑原子的散射时，可以忽略原子核的散射，而仅仅考虑电子的散射。如果忽略电子在原子中的分布，而把原子中的所有电子看成集中在一点上的点原子，那么某一种原子的散射能力就等于原子序数(Z)。因此一个原子的散射能力可用一个电子的散射能力的 f 倍表示。f 称为原子的散射因子。

实际上原子的电子并不集中在原子的中心，而是分布在直径与 λ 相当的范围内，从而各电子的散射波之间会相互干涉，一般 f 是 $\sin\theta/\lambda$ 的函数，θ 为布拉格角，λ 为波长，f 值随 θ 的增

加而下降；当 θ 固定时，f 随 λ 的变短而增加；当 $\theta=0$ 时，$f=Z$，图3.22示出了若干种原子的散射因子。

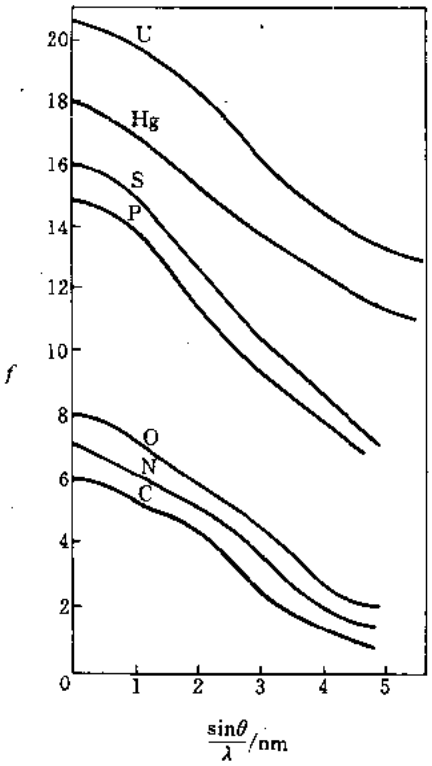


图3.22 原子散射因子曲线

另外，原子并不是处在静止状态，而是存在着热运动。由于这种热运动而降低了原子散射波的强度，这种强度的降低是按一个称为温度因子 $e^{-B(\sin^2\theta/\lambda^2)}$ 数值降低，$B=8\pi^2\overline{u}^2/3$，$\overline{u}$ 是原子沿垂直于反射面方向的平均位移。在蛋白质晶体中 $B=1.2-2\text{nm}^2$，相当于平均位移 0.015—0.05nm 之间。B 值越大，则衍射强度随 θ 角的增大而下降越快。B 由经验决定，实际考虑了多种因素，如晶体结构的某些统计无序性、分子的柔性、强度测量的误差以及晶体对射线的吸收等，因此平均位移值仅仅是一个粗略的近似。由于原子的振动并非各向同性，因此 B 值一般应作各向异性处理。

3.2.2 晶胞的散射强度

晶胞的内容是由生物大分子构成，而分子由各种原子组成。某一级衍射 hkl 是由晶胞内每一个原子散射的次级波所叠加而成。如果晶胞内有 N 个原子(图3.23)，在某一 hkl 衍射中，合成电磁波的电场强度将由 N 个原子的次生电磁波所合成。假设在晶胞原点的原子衍射波的周相角为零；而 N 个原子散射波的振幅为 $f_1,f_2,\cdots,f_N$；周相角为 $\varphi_1,\varphi_2,\cdots,\varphi_N$，则每个原子的散射因子：

$$f_1=f_1e^{i\varphi_1},f_2=f_2e^{i\varphi_2},\cdots,f_N=f_Ne^{i\varphi_N}$$

因此 N 个原子的次生波的合成波：

$$F_{hkl}=f_1+f_2+\cdots+f_N=\sum_{j=1}^{N}f_je^{i\varphi_j} \tag{1}$$

而 $\varphi_j=2\pi\frac{\Delta_j}{\lambda}$，其中 Δ_j 为第 j 个原子与晶胞原点的光程差。则 $\Delta_j=r_j\cdot(S-S_0)$，r_j 为第 j 个原子离晶胞原点的距离，而 $r_j=x_ja+y_jb+z_jc$，a,b,c 为晶胞的边长，x_j,y_j,z_j 为分数坐标，代入上式得：

$$\begin{aligned}\varphi_j&=2\pi(x_ja+y_jb+z_jc)\cdot(s-s_0)/\lambda\\&=2\pi/\lambda(x_ja\cdot(s-s_0)+y_jb\cdot(s-s_0)\\&\quad+z_jc\cdot(s-s_0)\end{aligned}$$

根据劳厄方程，式中 $a\cdot(s-s_0)=h\lambda,b\cdot(s-s_0)=k\lambda,c\cdot(s-s_0)=l\lambda$，代入上式得：

$$\varphi_j=2\pi(hx_j+ky_j+lz_j)$$

代入式(1)得：

$$F_{hkl}=\sum_{j=1}^{N}f_je^{i2\pi(hx_j+ky_j+lz_j)}$$

F_{hkl} 称为衍射 hkl 的结构因子，因为 F_{hkl} 与结构中的不同种类原子的 f_j 和原子的数量、分布 (x_j,y_j,z_j) 有关，即与分子结构有关。

F_{hkl} 可改写为：

$$F_{hkl} = |F_{hkl}|e^{i\varphi_{hkl}}$$

$|F_{hkl}|$称为衍射 hkl 的结构振幅，φ_{hkl}为衍射 hkl 的位相。

结构因子是一个复数，可在复平面上表示。图3.24示出晶胞内6个原子在 hkl 衍射中的结构因子 F_{hkl}。

在X射线衍射实验中，可用X光胶片或者X射线计数管把每个衍射强度记录下来，而强度正比于结构振幅的平方值，即

$$I_{hkl} \propto |F_{hkl}|^2$$

但是，不可能把每一个衍射点的位相都记录下来，因而不能从强度直接推算结构因子，而只能从每个强度推算结构振幅，这给进一步的结构分析工作带来了困难，不可能由衍射强度数据直接计算原子在晶胞中的分布。

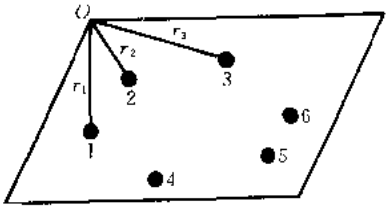


图3.23 晶胞中原子的分布

3.2.3 晶体的衍射强度

晶胞是组成晶体的基本单位，因此整个晶体产生的每个衍射线是由每个晶胞散射的X射线所叠加而成，例如由 T 个晶胞组成的一维晶体，假如重复周期为 a，那末如果在原点的第一个晶胞的散射波为 F_{hkl}，则第二个晶胞相对于同样的原点的散射波应为 $F_{hkl}e^{2\pi i a\cdot(s-s_0)}$，第 N 个晶胞相对于原点的散射波应是 $F_{hkl}e^{2\pi i(n-1)a\cdot(s-s_0)}$。因此整个晶体的散射波是：

$$G_{hkl} = \sum_{n=1}^{T} F_{hkl}e^{2\pi i(n-1)a\cdot(s-s_0)}$$

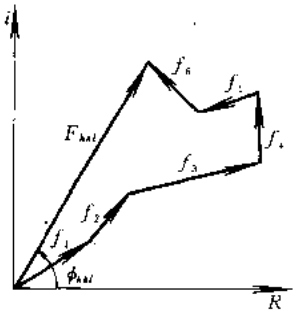


图3.24 结构因子的向量表示

当位相差 $2\pi a\cdot(s-s_0)=2\pi h$ 时，式中 h 为任意整数，则每个晶胞的散射将相互加强，它的结构振幅的大小正比于 $T\times|G_{hkl}|$，即衍射强度 $I_{hkl}\doteq T\times|G_{hkl}|^2$。所以一般说来，衍射强度正比于晶体的体积。因此，适合于X射线衍射分析用的晶体需要一定的大小。

3.2.4 衍射花样中强度分布的中心对称定律

如前所述晶体产生的每一个衍射可以看成是由晶体内相应的一组平面所产生，而每一组平面由确定的、一定种类的原子构成，因此产生的衍射线有固定的强度，同时由于每一组通过原点的平面(hkl)能产生两个强度相等的衍射点 I_{hkl}和 $I_{\bar{h}\bar{k}\bar{l}}$，它们分别由平面组$(hkl)$和$(\bar{h}\bar{k}\bar{l})$产生(图3.25)。同理可推知晶体内任何一组平面，都能产生两个强度相等的衍射点，因此不论晶体结构内是否存在对称中心，所产生的衍射花样的强度分布总是中心对称的。这一规律称为中心对称定律，或称为弗里德耳(Friedel)定律。虽然在晶体结构中可能存在32个点群，但是由于晶体的衍射强度分布的中心对称性，因此在衍射花样中可能存在的点群，仅为具有对称中心的11个点群，称为11个劳厄群(表3.2)。

表3.2 11个劳厄群

三斜	单斜	四方	三方	六方	正交	立方
$\bar{1}$	$2/m$	$4/m$	$\bar{3}$	$6/m$	mmm	$m3$
		$4/mmm$	$\bar{3}m$	$6/mmm$		$m3m$

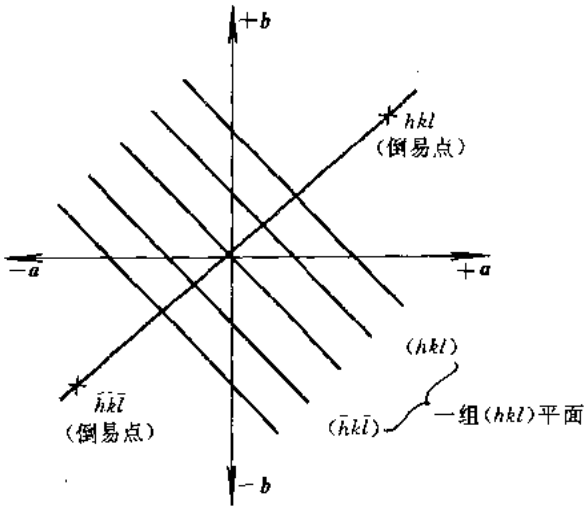


图3.25　倒易点 hkl 和 $\bar{h}\bar{k}\bar{l}$

3.2.5　晶胞中原子的数量对强度的影响

在前面讨论晶胞的散射强度时，已提到结构因子与晶胞中原子的种类、数量和分布有关，即结构因子与分子的结构有关，因为晶胞中不同种类的原子、数量和分布决定于分子的化学和立体的结构，而衍射花样中每个衍射点的强度取决于每个衍射点的结构因子的振幅值，因此整个衍射花样的强度分布特点决定于晶体的分子结构。在此为了着重讨论晶胞中原子数量仅影响衍射线的强度而不改变衍射方向的问题，我们以由两个原子组成的结构基元，按空间格子结点分布的结构，即每个晶胞中仅包含两个原子的结构为例(图3.26)。

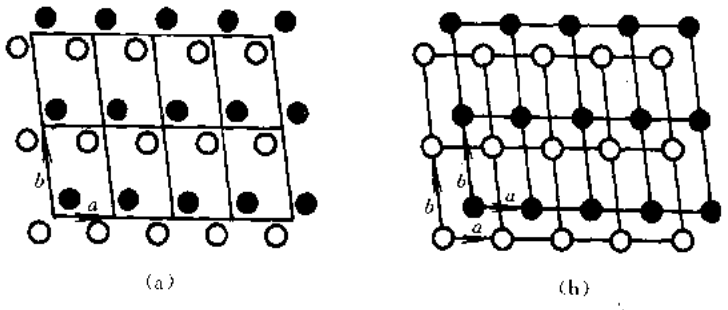


图　3.26

(a)结构基元为两个原子的结构；(b)分解成结构基元为一个原子的两个结构

对于上述这样的结构，在形式上总可以将它分解成两个相等且彼此平行，并穿插在一起的单原子结构，它们的格子型式以及晶胞大小、形状完全一样，与没有分解前的两个原子结构的空间格子型式，以及晶胞大小和形状亦完全一样(图3.26(b))。正由于分解出来的单原子晶格彼此间相互平行又具有完全相同的晶胞大小和形状，因此产生衍射的方向与没有分解前的结构是完全一致的。由此推而广之，对具有同样晶胞大小和形状的结构来说，晶胞内原子数量的增加，并不导致新的衍射方向。但是衍射的强度则由于晶胞内原子数量的增加而引起变化。这是由于所分解出来的各个结构的"反射面"，虽然都是相互平行的，但一般并不重叠在一起，而是有一定距离。结果，由不同性质的、不同平移向量联系起来的原子平面所"反射"的射线之间，将出现光程差，而且这一光程差一般不等于波长的整数倍，因而相互干涉就会使衍射强度发生变化。由于任何一个具有复杂结构，如生物大分子结构，所分解出来的、具有单原子结构的数目总是有限的，因此，除特殊情况外，干涉的结果一般并不会导致衍射强度的消失，而仅是强度有所改变，其改变程度取决于各"反射"线振幅之大小，以及相互间的光程差，即取决于组成各"反

射”面的原子性质及各种面之间的距离。图3.27(a)示意了分解出来的两个不同面之间“反射”线相互干涉对强度的影响。假如在100“反射”中，相距为 d 的两个平面间的光程差为1λ，而不同原子面之间的距离，假如为 $d/3$，则两个不同面间的“反射”线的光程差为 $\lambda/3$，按公式 $2d\sin\theta=n\lambda$，n 不为整数，各“反射”线间相互干涉，而使强度减弱[图3.27(b)]。

晶体衍射线的强度，一般地讲，反比于晶胞内原子的数量。数量多则衍射强度弱。因此生物大分子的分子量越高，原子数越多，构成晶体的晶胞将越大、衍射强度将越弱。为收集更强的衍射强度，就要求培养出更大的分析用的晶体，这给生物大分子的晶体分析工作带来了困难。反之两个包含完全相同数量和种类原子的晶胞，仅原子在晶胞中分布不同的晶体，则将出现不同的衍射强度分布，这说明分子的立体结构信息存在于衍射强度之中。

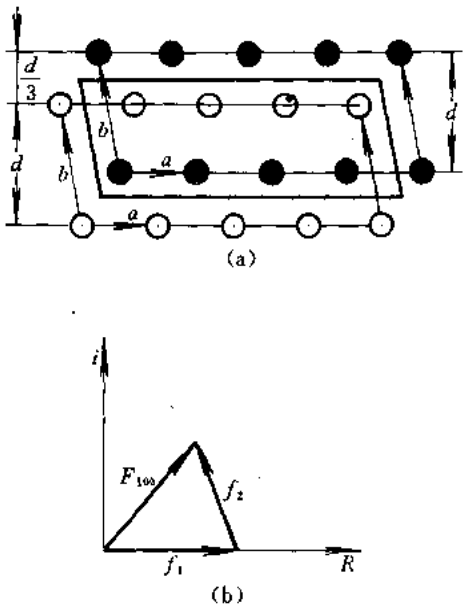


图 3.27

(a)示意两个不同平面组之间的距离差；

(b)f_1、f_2分别表示不同面组的散射因子

F_{100}表示合成的散射因子

3.2.6 空间格子、对称元素对衍射强度的影响

3.2.6.1 不同格子型式对强度的影响

在前面讨论劳厄衍射条件时，都是以简单格子为基础的，而实际晶体中却存在着复格子。图3.28(a)示出了两种不同型式的立方格子在平面上的投影，假如在每个格子结点上都由同一种原子构成。图(a)上部表示一个平面正方简单格子结构，晶胞中仅有一个原子，假定坐标为(x,y)，那么在100，200…等各级衍射中，根据公式

$$F_{hkl}=\sum_{j=1}^{N} f_j e^{i2\pi(hx+ky)},$$

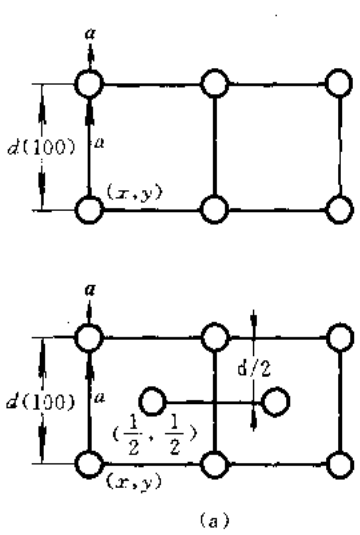
则 $F_{100}=fe^{i2\pi x}$，$F_{200}=fe^{i4\pi x}$，…，$F_{h00}=f_1$，因而各级衍射强度(I_{h00})都相等。同理，根据布拉格公式，当 $n=1,2$…等各级“反射”中，在(100)平面组中相邻面间的“反射”线的光程差为1λ，2λ，…，$n\lambda$，都为波长的整数倍，“反射”线相互加强[图3.28(b)上]。而对于一个平面正方体心格子结构[图3.28(a)下]来说，晶胞中有两个原子 A 和 B，假定坐标分别为(x,y)和$(x+1/2,y+1/2)$，则根据公式：

$$\begin{aligned}F_{hk0}&=f_1e^{i2\pi(hx+ky)}+f_2e^{i2\pi[h(x+1/2)+k(y+1/2)]}\\&=fe^{i2\pi(hx+ky)}+f_2e^{i2\pi(hx+ky)}\cdot e^{i2\pi(h/2+k/2)}\\&=fe^{i2\pi(hx+ky)}(1+e^{i\pi(h+k)})\end{aligned}$$

当 $h+k=$奇数时，$F_{hk0}=0$

当 $h+k=$偶数时，$F_{hk0}=2fe^{i2\pi(hx+ky)}$

由此在晶体的 $hk0$衍射花样中，凡是 $h+k=$奇数指标的衍射强度都为零(系统消失)。同理可按布拉格公式分析在 $n=1$的“反射”中，平面组(100)中相邻面间的光程差为1λ，但是与相邻的平面组(200)之间“反射”线的光程差为 $\lambda/2$，由于两组平面都由同种原子组成，因而相干的结果，使该方向强度为零。在 $n=2$的“反射”中，不同平面组间的“反射”线的光程差为1λ，因而“反射”线相互加强[图3.28(b)右]



f_2 $\varphi_2=2\pi$ 100反射： $F_{100}=2f_1$

f_1 $\varphi_1=0$ $\vec{f_1}$ $\vec{f_2}$

200反射： $F_{200}=2f_1$

f_2 $\varphi_2=2\pi$ 100反射： $F_{100}=0$

f_3 $\varphi_3=\pi$ $f_1 \longrightarrow$ $\longleftarrow f_3$

f_1 $\varphi_1=0$ 200反射： $F_{200}=2f_1$

$\vec{f_1}$ $\vec{f_2}$

(b)

图 3.28

(a) 上为正方简单格子，下为正方平面体心格子；(b) f_1，f_2 分别代表由平移量 a 所联系的相邻面的散射因子；φ_1、φ_2 为 f_1，f_2 的位相；f_3，φ_3 分别代表由 B 原子构成的平面的散射因子和位相

同理，可推广至三维体心格子的消光规律，$h+k+l=$奇数的衍射点强度为零。其他非简单格子的消光规律，可用同样的原理和方法推引。

3.2.6.2 螺旋轴、滑移面对衍射强度的影响

在晶体结构中若存在某种螺旋轴或滑移面，则一定存在由平移向量联系起来的分子或原子团。这样的原子排列对X射线衍射强度会产生系统的影响，例如在 c 方向存在一个二重螺旋轴(图3.29)，则若在晶胞中存在一个原子 A(x,y,z)，必有2_1联系起来的原子 B$(\bar{x},\bar{y},z+1/2)$，则

$$F_{hkl}=f_A e^{i2\pi(hx+ky+lz)}+f_B e^{i2\pi[h\bar{x}+k\bar{y}+l(z+\frac{1}{2})]}$$

$$=f_A e^{i2\pi(hx+ky+lz)}+f_B e^{i2\pi(h\bar{x}+k\bar{y}+lz)}\cdot e^{inl}$$

在$00l$衍射中，

$$F_{00l}=f_A e^{i2\pi lz}+f_B e^{i2\pi lz}\cdot e^{inl}=f_A e^{i2\pi l}(1+e^{inl})$$

当 $l=$奇数时，$F_{00l}=0$；

当 $l=$偶数时，$F_{00l}=2f_A e^{i2\pi l}$

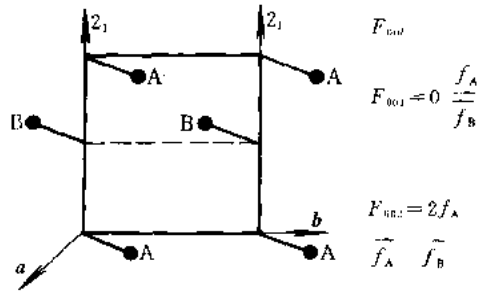


图3.29 2_1螺旋轴联系的原子和 F_{001}，F_{002} 的向量合成

同理根据布拉格方程，在1级衍射中，由图若原子平面 AA 之间的光程差为1λ，则 AA 与 BB 面间的光程差为$1/2\lambda$，"反射"线相互干涉而使强度为零。而在2级衍射中，AA 与 BB 之间的光程差为1λ，相互干涉而使强度加强。以此类推，当 n 为奇数的衍射点将系统地消失。

如果晶体结构中存在某种滑移面，由于该对称元素亦存在一个平移量，因此必然引起系统消光。例如在单斜晶系中存在一个垂直于 b 轴的 a 滑移面，若在晶胞中有一个原子 A(x,y,z)，必有另一个原子 B$(x+1/2,\bar{y},z)$(图3.30)，则

$$F_{hkl}=f_1 e^{i2\pi(hx+ky+lz)}+f_2 e^{i2\pi(hx+h/2-ky+lz)}$$

在 $h0l$ 衍射中，

$$F_{h0l} = fe^{i2\pi(hx+lz)}(1 + e^{i\pi h})$$

当 h＝奇数时，$F_{h0l}=0$，h＝偶数时，$F_{h0l}=2f$。

据布拉格方程，在1级衍射中，或 n＝奇数时，A 原子平面与 B 原子平面间的光程差为1/2λ，各次级散射波相互干涉，衍射点强度系统消光而为零。当 n＝偶数时，A 与 B 两原子平面间的光程差为 λ 的整数倍，各波相互干涉而系统加强。

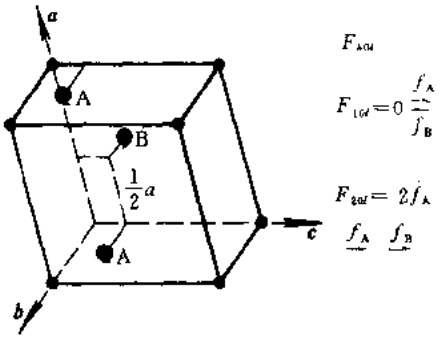


图3.30 a 滑移面联系的原子和 F_{1ol}，F_{2ol}的向量合成

由格子的不同型式、螺旋轴和滑移面所引起的衍射点的强度的消失是有规律的，称为系统消光。据此通过分析衍射花样中强度的系统消光的特点，可判断晶体的晶格型式，以及是否存在某种螺旋轴或某种滑移面等微观对称元素，从而可以测定晶体存在的或可能存在的空间群。表3.3和表3.4列出了系统消光规律与空间格子和对称元素间的关系[14]。

表3.3 空间格子的消光规律

空间格子类型	消光规律
体心	hkl 反射中 $h+k+l=2n+1$
A 面心(100面)	hkl 反射中 $k+l=2n+1$
B 面心(010面)	hkl 反射中 $h+l=2n+1$
C 面心(001面)	hkl 反射中 $h+k=2n+1$
面心	hkl 反射中 $h+k, k+l, l+h=2n+1$

表3.4 在蛋白质晶体结构中存在的对称元素的消光规律

对称元素	对称元素方向	反射类型	消光规律
二重螺旋轴(2_1) 四重螺旋轴(4_2) 六重螺旋轴(6_3)　三者消光规律相同	a	$h00$	$h=2n+1$
	b	$0k0$	$k=2n+1$
	c	$00l$	$l=2n+1$
三重螺旋轴(3_1,3_2) 六重螺旋轴(6_2,6_4)	c	$00l$	$l=3n+1, 3n+2$
四重螺旋轴(4_1,4_3)	a	$h00$	$h=4n+1$，2或3
	b	$0k0$	$k=4n+1$，2或3
	c	$00l$	$l=4n+1$，2或3
六重螺旋轴(6_1,6_5)	c	$00l$	$l=6n+1$，2，3，4或5

第四章　记录衍射花样的方法

记录衍射花样(衍射线的几何分布和强度)的目的,第一在于通过测量衍射点的分布计算倒易晶胞的大小、形状,由此算出晶胞的大小和形状,同时通过分析衍射花样的对称性,测定晶体所属点群,以及根据系统消光规律确定晶格的型式和可能的空间群等;第二测量每个衍射点 hkl 的强度,并由强度计算出结构因子的振幅值,为确定晶胞中的原子分布准备一套衍射数据。

记录衍射花样的方法,随着晶体结构分析工作的发展亦不断改进,特别是电子计算机的应用,使 X 射线衍射数据的收集和处理更加精确、高速、自动化。但是根据记录衍射花样的不同方式可分两大类:胶片法和计数管法。胶片法主要是用 X 光胶片记录衍射花样,使用的方法主要有回摆法、旋进法、魏森堡法等;计数管法主要用正比计数管记录衍射花样,主要使用的仪器为四圆衍射仪、面积探测仪等。这些专门的仪器结构较复杂、有专著介绍,在此仅为进一步理解 X 射线衍射法的基本原理,主要介绍回摆照相法、旋进照相法、魏森堡照相法和四圆衍射仪的基本原理。

4.1　X 射线的产生和一般性质

一般产生 X 射线都是借助阴极射线(高速电子流)对物质的碰撞来获得。

X 射线机上用的封闭式 X 射线管(图4.1)为一高度真空管,管内主要由阴极和阳极(靶)

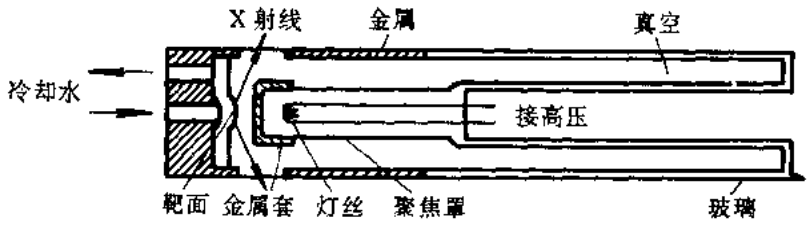


图4.1　封闭式 X 射线管示意图

两部分组成,阴极为钨丝。当通电时,钨丝炽热而在真空中放出自由电子,此时若在两极间加以高电压(30—60kV),则在此电场的作用下,阴极附近的自由电子以接近光速的速度向阳极作加速运动,当它们到达阳极靶面时突然被阻止,其一部分动能即转变而产生 X 射线,而绝大部分能量转变成热能。因此阳极要很好地通水冷却,否则很快就把 X 射线管烧坏。

由 X 光管产生的 X 射线包括"白色"X 射线和特征 X 射线两部分。白色射线包括了一系列波长的连续光谱,它是由于高速电子与阳极撞击时电子被阻止的方式不同,有的在靶面停止,有的穿透到靶面层,深浅不一,失去能量多少不同,因而发射的 X 射线波长亦不同,形成一个连续光谱。"白色"X 射线的性质与加在 X 射线管两极的电压有关,与靶材料无关。假如所加电

压为 V，将产生的“白色”X 射线的最短波长为 λ_0，则

$$\lambda_0 = \frac{hc}{eV} = \frac{12398}{V}$$

连续波的强度随波长而改变(图4.2)。

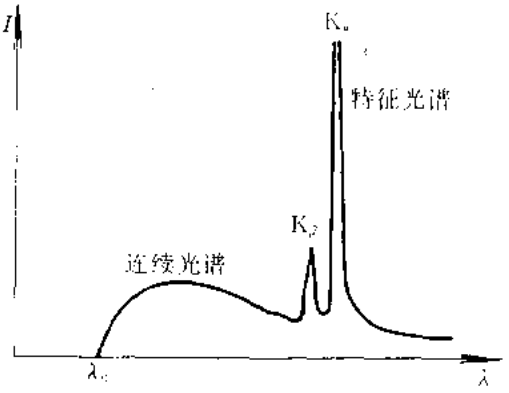


图4.2 铜靶 X 射线光谱

特征射线是由阳极材料决定的有特定波长的 X 射线。特征 X 射线的产生是由于高速电子把阳极靶原子内层如 K 层的电子赶出，再由外层电子跃迁进去，由于位能的下降而发出的 X 射线。一种靶材料能被同时激发出若干条波长不同的特征谱线，如铜靶在36kV 电压下，能发射出 $K_{\alpha1}$，$K_{\alpha2}$，K_β 等射线。晶体结构分析主要用强度最大的 K_α 射线，其他谱线可设法(如用晶体单色器)滤去。

X 射线与可见光、无线电波一样，为电磁波。它直线传播，波长较短，在晶体结构分析中所用的波长一般在0.05—0.25nm 范围内。

X 射线折射率很小，穿透力强，能被物质吸收，因此在使用时必须注意防护和安全操作。晶体衍射实验用的 X 射线较医学上用作人体透视的 X 射线波长要长，更易被人体吸收，所以危害更大。但是衍射所用的 X 射线束较集中，易于防护，只要注意安全操作，是可以防止对机体的损伤的。

4.2 蛋白质晶体的挑选和安装

生物大分子晶体，例如蛋白质晶体，不同于一般的有机分子晶体，含有大量的溶剂分子，可高达40%左右。由此晶体包含着两相，一是固相，由蛋白质分子，通过分子间的几个接触点形成晶格，而另一相是充满晶格的液相。直接与蛋白质分子接触的溶剂，通常是有序的，能同分子表面的极性基团形成很强的氢键，而大量溶剂分子是无序的，这些溶剂分子对维持整个晶体结构的有序性相当重要，一旦晶体暴露在空气中，溶剂将蒸发而使蛋白质分子堆积更紧密，并失去原有的分子排列的规则性。最后晶体失去衍射能力而无法应用。因此在进行 X 射线衍射实验以前，必须仔细挑选晶体，然后把晶体安装在毛细管内，设法保持在一个高湿度的环境里。

由于生物大分子的分子量很高，由大量轻原子组成，结构复杂，因而形成的晶胞的体积都较大，再加上溶剂含量较多，因此晶体的衍射本领较弱。为此在挑选晶体时，应综合考虑分子量的大小、光源的强弱等因素，要在偏光显微镜下仔细观察和挑选无裂缝、透明度好、线度清晰的单晶，大小在0.8×0.6×0.4mm^{3}左右。然后根据已挑选晶体的大小，选择管壁很薄(0.001mm)直径略大于晶体的玻璃管或石英管。玻璃管要求硬质，不含重原子，因重原子能吸收和散射 X 射线。

把晶体引进到较细的管中是一件细致、要有耐性和经验的工作，稍不注意就很容易把长期辛苦培养的晶体毁于一旦。所使用的方法较多，有利用毛细管的虹吸作用把晶体吸入的，亦有用注射器连接细管把晶体吸入的等等，可根据各人的实践和经验选择。在装进晶体以后，根据照相的要求设法调整晶体在管中位置和取向，然后小心吸去晶体周围过多的溶剂，并在管两端

加一小滴结晶母液，最后用真空蜡把两端密封(图4.3)[10~16]。

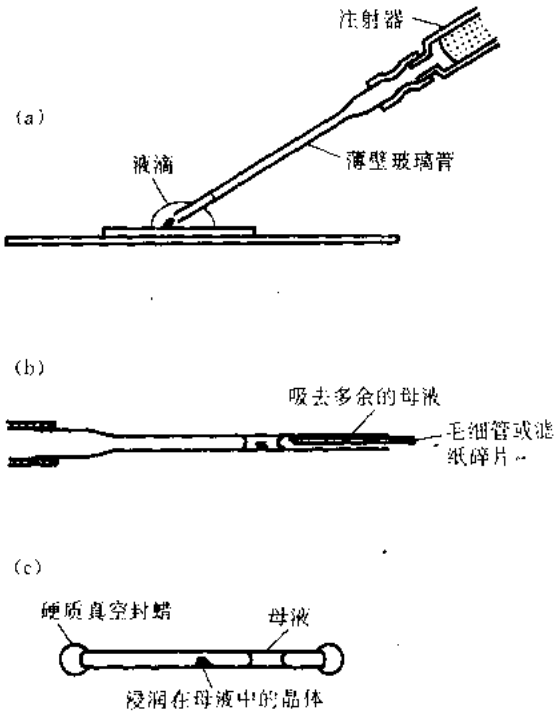


图4.3 装晶体的步骤
(a)吸晶体至玻璃管中；(b)吸去晶体附近的多余液体；(c)已装好的晶体

4.3 回摆照相法

回摆照相是采用单色X射线，晶体装在测角头上，晶体的取向与摆动轴平行。为了保证有足够的衍射产生，晶体绕轴来回摆动，摆动角度可根据实验需要确定。在离晶体的一定距离处安置一张圆筒形或平板底片，入射X射线垂直于摆动轴(图4.4左)。

按劳厄方程，平行于摆动轴的每一直线原子列所产生的衍射效应，将形成不同级的以摆轴为轴的一系列圆锥面。这些圆锥与圆筒形底片相交均为一个圆。如果将底片展平，则展现出一系列平行线。但是，晶体是三维空间结构，因此所产生的衍射效应，除必须服从轴方向的劳厄条件以外，还必须服从另两个方向的劳厄条件，从而产生的衍射线将出现在各衍射圆锥的相截线上，而形成某些不连续的线，它们与圆筒形底片相截，则在底片上产生相应的衍射点。当底片展开时，即成为分布在若干条平

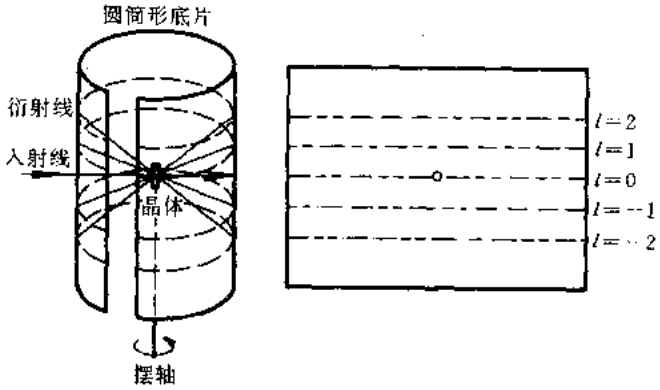


图4.4 回摆照相装置原理和底片展开后的层线

行线上的斑点(图4.4右)。

在回摆图上，衍射点分布的直线称为层线。通过底片中心的层线称为0层线，相当于$l=0$的圆锥在底片上的相截点。从0层向上，依次为+1层、+2层…等，从0层往下依次为-1层、-2

层…等。它们分别与 $l=\pm1, l=\pm2$…等圆锥在底片上的相截点相对应。

回摆法可用于调整晶体定向、观察晶体对称性和分辨率。在测定生物纤维状物质(详见4.7.1)的周期或晶胞参数时，回摆图上，层线的距离与晶体摆动轴的周期存在着一定关系。假如平行于摆轴的晶轴的周期为 c，晶轴底片的距离，即圆筒形底片的半径为 R，第 l 层至0层之间的距离为 y_l(图4.5)。由

$$c\cos r_l = l\lambda$$

得

$$c = \frac{l\lambda}{\cos r_l}$$

而

$$\cos r_l = \frac{y_l}{\sqrt{R^2 + y_l^2}}$$

则

$$c = l\lambda\sqrt{1 + (R/y_l)^2}$$

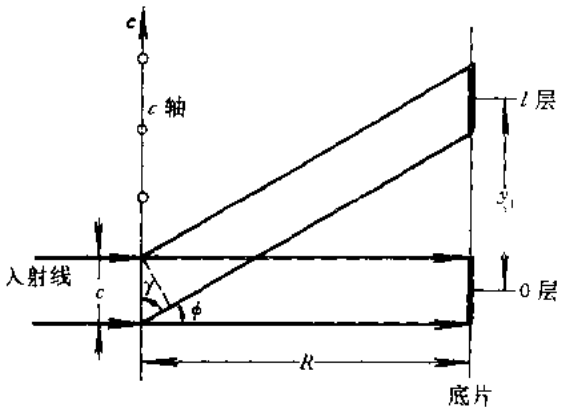


图4.5 层线间距与 c 方向 c 的关系

圆筒形底片之半径 R 可实际测定，λ 为已知波长，l 为所测层线的级别，y_l 可从底片上测得，代入上式就可求出 c 轴的单位长度。其他两个方向 b 和 a 的晶胞轴长亦可按同样方法摄取 b 方向和 a 方向的回摆图求得。

回摆法在生物大分子单晶衍射数据收集中主要应用在以潜像板记录衍射数据的面积探测仪上。这时，晶体的定向不必一定要与摆轴平行、摆动角度按晶胞大小确定，一般在2°左右。在像板上记录的每幅衍射花样，经仪器自动地读出后，在屏幕上显示，并把数据储存在计算机内。

4.4 旋进照相法

旋进相机是一种旋进晶体，旋进底片的单晶相机，它能拍摄不歪曲的、所选定的倒易格子层。它是蛋白质晶体结构分析最基本的工具，用它测定晶体的对称性、空间群以及检查重原子衍生物的强度变化等，在很多情况下亦用来收集强度数据。这种相机在1942年由贝尔格尔(Buerger)设计，故称贝尔格尔旋进相机(图4.6)。

4.4.1 基本原理

设想一个倒易格子平面 POQ 垂直于入射X射线 CO，并与反射球相切于 O 点，其法线 ON 与 CO 重合，如在倒易格子平面后放置一张平板底片，仅能记录入射线(图4.7)。假若使倒易格子平面的法线 ON 倾斜一个 μ 角，这时倒易格子与反射球相交成一个圆。如果在与倒易格子平面相平行的位置放一张平板底片，就可记录到衍射花样(图4.8)，衍射点都分布在 $P'O'$ 为直径的圆周上。现在将 μ 角保持恒定，并使 ON 法线绕着X射线旋进，从 $X\rightarrow X'\rightarrow X$(图4.9)，使倒易格子平面的各倒易格子点随着旋进运动与反射球不断相碰，满足产生衍射的条件，图4.10示意记录到的几个与反射球相碰时的特殊位置。

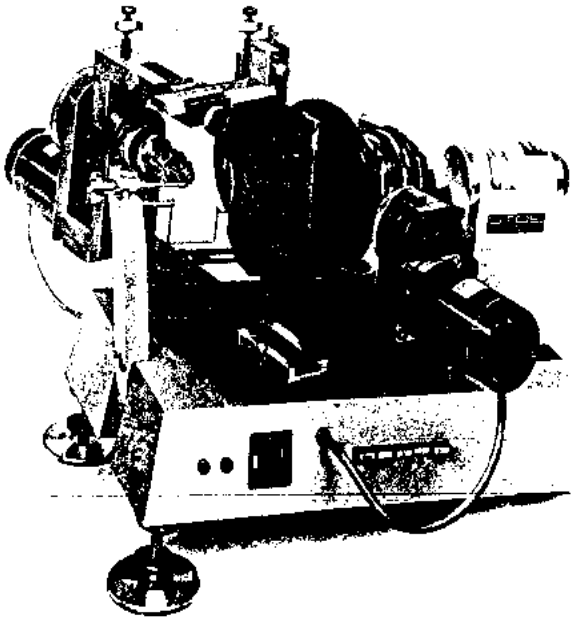


图4.6　旋进相机

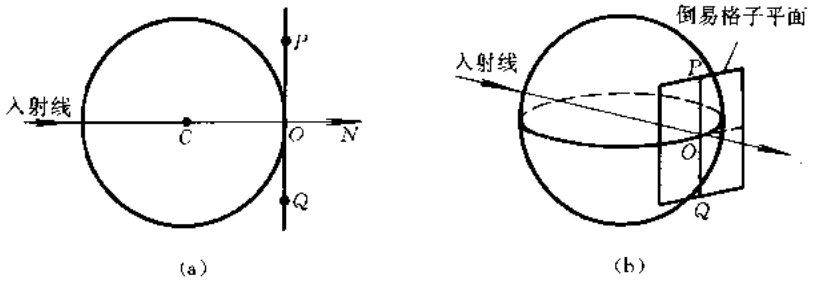


图4.7　倒易格子平面垂直于入射线，并与反射球相切
(a)剖面图；(b)立体图

如果底片始终保持平行于 POQ 平面，并随着该平面同步旋进，当各倒易格子点与反射球相碰时，衍射产生，底片就可记录到各个“反射”线。这样记录的全部衍射花样，将与倒易格子层的类型完全一致。

4.4.2　各层倒易格子平面的分离

在上节只讨论了一层倒易格子与反射球相碰的过程，实际上在旋进过程中各层倒易格子都不同程度地与反射球相碰。因此要保证每次实验仅记录到所需的一层倒易格子，就必须插入一个称为层线屏的装置，它只允许某一层的衍射圆锥通过，而屏蔽了其他各层的衍射圆锥。

4.4.2.1　零层层线屏的位置

由图4.11可知，零层层线屏至晶体的距离 S_0 与旋进角 μ 的关系：

$$\cot\mu = \frac{S_0}{r_s}$$

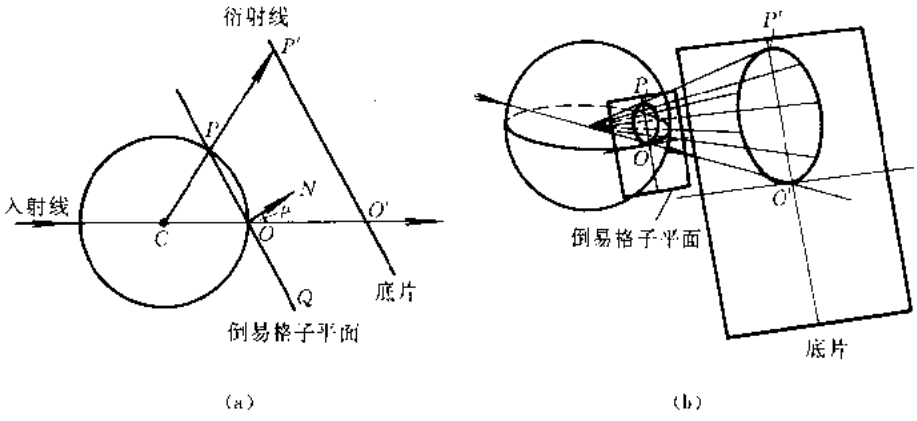


图4.8 倒易格子倾斜一个 μ 角后，与反射球相截
(a) 剖面图；(b) 立体图

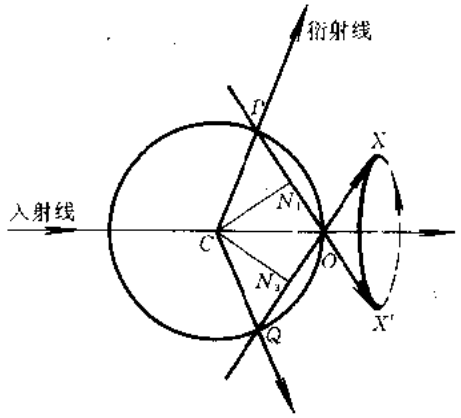


图4.9 倒易格子平面的法线绕入射线旋进

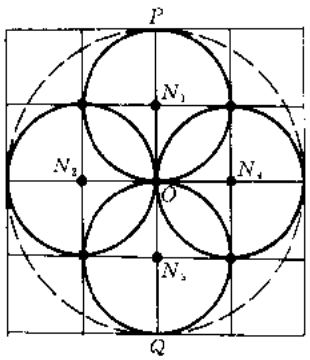


图4.10 在旋进过程中倒易格子平面与反射球相碰时的4个特殊位置

则 $$S_0 = r_s\cot\mu$$

r_s 为层线屏半径。由实验所要求的 μ 值和选定的 r_s 值即可由上式求 S_0 值。

4.4.2.2 非零层层线屏的位置

由图4.12得

$$\cot\nu_n = \frac{S_n}{r_s}$$

则 $$S_n = \cot\nu_n r_s$$

ν_n 为第 n 层衍射圆锥半顶角，S_n 为第 n 层层线屏至晶体的距离。

ν_n 由校正晶体时的静止照片上求得：

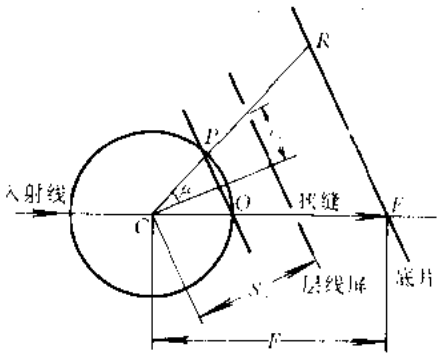


图4.11 零层层线屏的位置

$$\tan\nu_n = \frac{r_n}{F}$$

r_n 为第 n 层至底片上入射线中心距离，F 为晶体至底片距离。

或由 $$\zeta = \cos\mu - \cos\nu_n$$

$$\cos\nu_n = \cos\mu - \zeta$$

$$\nu_n = \cos^{-1}(\cos\mu - \zeta)$$

ζ 为第 n 层倒易格子平面至零层的距离。ζ 由实验求得，μ 由实验要求确定。

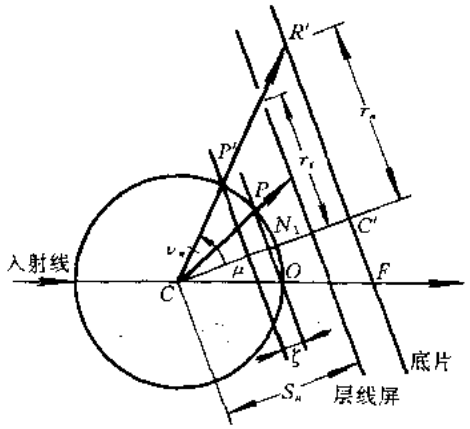


图4.12 非零层层线屏的位置

4.4.3 晶体方向的校准

要分别记录各层的衍射花样，必须通过层线屏把各层衍射圆锥分开，这就要求晶体的定向必须很准确(偏差小于5′)，否则各层不但不能分开，而且一层亦不可能记录完全。为此在安装晶体时就应该尽量按预定方向装准，在此基础上再通过静止及小角度旋进照相进一步把晶体方向校准。

4.4.3.1 静止照相校正方向

静止照相是 $\mu=0$，晶体不摆动时，摄取衍射花样。假如晶体的定向已较准，偏差小于20′，在这种状态下摄得的静止衍射花样，应该是各层倒易格子平面与反射球相截时产生的一系列分布在同心圆上的衍射点。圆心正好与入射线重合。

在实验中一般都会产生一定偏差，晶体定向不可能太准。这样零层倒易格子与反射球相切的点，就会偏离入射线方向。假如倒易格子平面向上方偏一个角度 ε_H(人面向入射线观察)，这

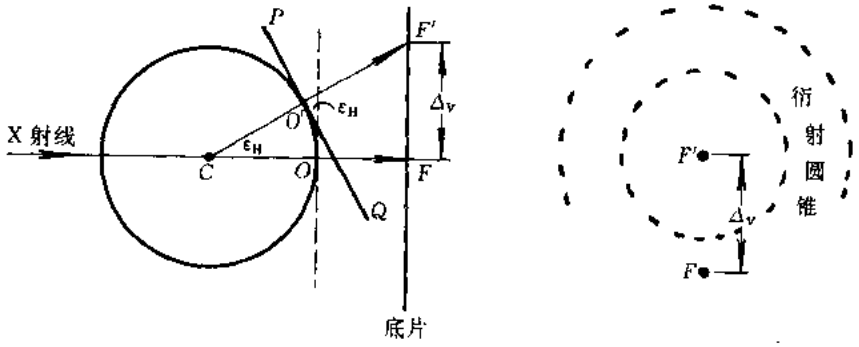


图 4.13

(a)零层倒易格子向上偏离时，与在底上记录到的衍射点之间的关系；(b)示意中心点偏离时的静止衍射图

时与反射球相切在 O' 点[图4.13(a)]，相应的衍射底片上亦将观察到向上偏离的衍射点 F'，离原点 F 的距离为 Δ_V[图4.13(b)]，由图中的几何关系可求得 ε_H。

$$\tan\varepsilon_H = \frac{\Delta_V}{\overline{CF}} = \frac{\Delta_V}{F}$$

CF 为晶体至底片的距离，以 F 表示。

然后通过调正测角头上的、相应的瓦板，校准晶体的方位。

根据同样的道理可以校准其他方向上的偏离。一般讲倒易格子平面在任何一个方向上的偏离，总可在相应的衍射底片上分解为水平方向(Δ_H)和垂直方向(Δ_V)的偏离。然后计算出相应的 ε_H 和 ε_V，再调正测角头上的水平瓦板和相机上 Φ 圆刻度盘。

4.4.3.2　小角度旋进照相校准方向

为进一步校准晶体在5′左右的偏离，需要通过 $\mu=3°$ 左右的小角度旋进照相。这种照相使零层倒易格子与反射球相碰在一个半径很小的范围内，从而可获得一部分的衍射花样。如果晶体定向足够正确，即可得到衍射图圆心与入射线重合的、衍射点分布对称的衍射图。如果晶体

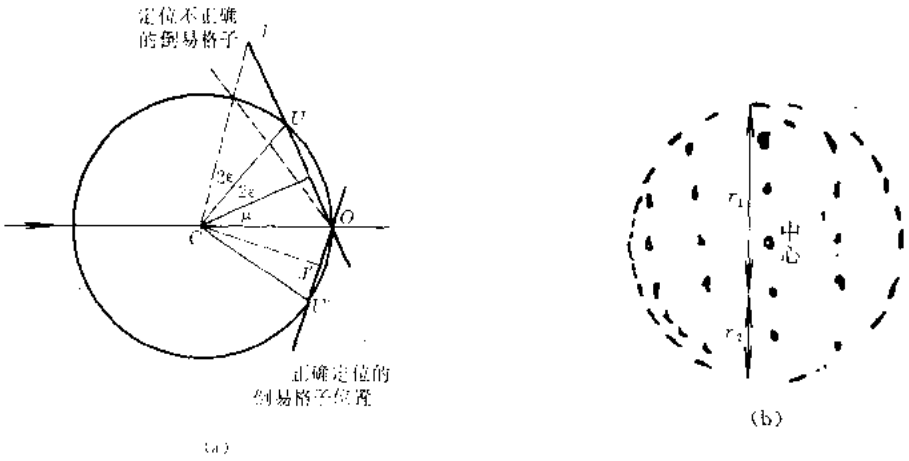


图　4.14

(a)倒易格子平面向上偏离时，小角度旋进照相中倒易格子平面与反射球相碰时的几何关系；(b)示意倒易格子平面向上偏离时的衍射花样

定向仍然不够准，例如向上偏离某一角度，那么零层倒易格子与反射球相碰的范围将产生上下不等的情况[图4.14(a)]，上部相截的半径(虚线)大于下半部，反映到衍射图上将使中心点偏向下方，r_1 代表偏离中心点至图边缘距离，r_2 代表偏离中心点至另一边的距离[图4.14(b)]。由衍射图可计算出：

$$\Delta = r_1 - r_2$$

由图4.14(a)和4.13(a)的几何关系可得：

$$r_1 = F(OJ) = F(OU + JU) = F\left[2\sin\mu + \frac{\sin 2\varepsilon}{\cos(2\varepsilon + \mu)}\right]$$

$$r_2 = F(OJ') = F(OU' - U'J') = F\left[2\sin\mu - \frac{\sin 2\varepsilon}{\cos(2\varepsilon - \mu)}\right]$$

$$\Delta = \frac{F2\sin 2\varepsilon\cos 2\varepsilon\cos\mu}{\cos^2 2\varepsilon\cos^2\mu - \sin^2 2\varepsilon\sin^2\mu}$$

假如 μ 和 ε 在很小的情况下，得

$$\Delta \doteq 4\varepsilon F$$

则

$$\varepsilon \doteq \Delta \cdot \frac{1}{4F}$$

F 为晶体至底片的距离(放大因子)。

同理可求得其他方向的偏离,然后调整测角头上的相应瓦板和相机上 Φ 圆刻度盘。

4.4.4 零层与非零层的旋进照相

经过小角度旋进照相校准方向后,一般晶体定向已达到较理想的程度,就可开始正式照相,即开始大角度,如 $\mu=15°$的旋进照相。

4.4.4.1 零层的旋进照相

第一,设置好层线屏。根据确定的 μ 角的大小和所用层线屏的半径,按4.4.2.1节公式计算出 S_0,然后按 S_0值装好层线屏。

第二,设置好旋进角 μ。

第三,检查层线屏的位置。

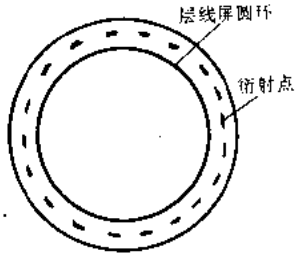


图4.15 示意通过层线屏狭缝的衍射花样

如果晶体定向和层线屏位置正确,零层衍射圆锥将通过层线屏狭缝而摄得如图4.15所示的圆锥的花样,证明一切位置正确,就可装好底片开始照相。

4.4.4.2 非零层旋进照相

基本步骤同零层旋进照相,但是,必须调整层线屏的距离和向靠近晶体方向移动底片盒,层线屏的距离 S_n 按4.4.2.2节公式计算。

在进行非零层的旋进照相时,必须移动底片盒,这是因为零层倒易格子 POQ 在作旋进运动时,上一层倒易格子的中心 O' 同时绕入射线作旋进运动,它的轨迹为 $x' \rightarrow x''$,如果底片仍放在 F 处,就不可能记录到没有歪曲的上一层的衍射花样(图4.16),所以必须把底片移至 F' 处。

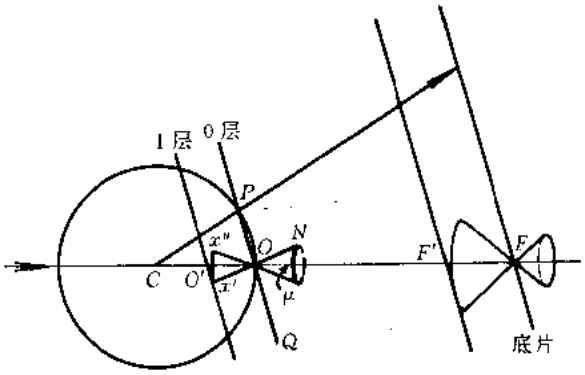


图4.16 非零层倒易格子在旋进过程中的旋进轨迹

设底片移动的距离为 S_f,则按图4.12所示,高层至零层的距离:

$$\zeta = \cos\mu - \cos\nu \qquad (1)$$

$$\tan\nu = \frac{R'C'}{CC'} = \frac{r_n}{F} \qquad (2)$$

r_n 为静止衍射图上零层至第 n 层间的距离,F 为底片至晶体间的距离。

把(2)式代入(1)式得:

$$\zeta = \cos\mu - \cos\left(\tan^{-1}\frac{r_n}{F}\right)$$

则

$$S_f = F\zeta$$

4.4.5 旋进照相法的应用

4.4.5.1 晶胞参数的测定

旋进照相提供了没有歪曲的倒易格子平面，所以很容易确定正确的倒易单位晶胞，并求出 $a^*,b^*,c^*,\alpha^*,\beta^*,\gamma^*$。例如在按 $\boldsymbol{a}^*\boldsymbol{b}^*$ 平面展开的零层衍射底片上，测得 $\boldsymbol{a}^*$ 方向两衍射点间的距离 t，则 $t=F\lambda a^*$，$a^*=t/F\lambda$，F 为放大因子，λ 为波长。同理可求得 b^*，以及 $\boldsymbol{a}^*$ 与 $\boldsymbol{b}^*$ 夹角 γ^*；由 $\boldsymbol{b}^*\boldsymbol{c}^*$，$\boldsymbol{a}^*\boldsymbol{c}^*$ 平面展开的衍射图可求得 $\boldsymbol{c}^*$ 和 $\boldsymbol{a}^*$，以及 $\boldsymbol{b}^*$ 与 $\boldsymbol{c}^*$ 夹角 α^*，$\boldsymbol{a}^*$ 与 $\boldsymbol{c}^*$ 夹角 β^*。

由倒易晶胞与实晶胞间的关系即可求得晶胞参数 $a,b,c,\alpha,\beta,\gamma$。例如正交晶系：

$$a = d_{(100)} = \frac{\lambda}{d^*_{100}} = \frac{1}{a^*}$$

$$b = d_{(010)} = \frac{\lambda}{d^*_{010}} = \frac{1}{b^*}$$

$$c = d_{(001)} = \frac{\lambda}{d^*_{001}} = \frac{1}{c^*}$$

$$\alpha = \alpha^*,\beta = \beta^*,\gamma = \gamma^*$$

4.4.5.2 空间群的测定

根据零层和高层衍射图上衍射点的系统消光或系统存在的规律，确定可能的空间格子类型(表3.3)，以及存在某些对称元素，或不存在某些对称元素(表3.4)。但是，在某些情况下不可能唯一地确定，例如3_1和3_2，4_1和4_3螺旋轴不能被区分，因此某些空间群不能唯一测定，有待进一步的工作。

测定空间群的步骤大体如下，例如斑头雁血红蛋白晶体的 $hk0$，$hk1$旋进图上存在$4mm$ 劳厄对称性，由此确定晶体属422点群；在 $hk0$照片上，$h00$型衍射中 $h=2n+1$消光，说明存在2_1对称元素；而 $hk1$照片上 $h+k+l=2n+1$无消光，表明空间格子类型为 P 格子，在$00l$ 中 $l=2n+1$消光，表明存在4_2对称元素。综合上述结果，测得晶体的空间群为 $P4_22_12$。在生物大分子晶体中可能出现的65个非对映空间群与衍射强度系统消光的关系见附录。[10]-[16]

4.4.5.3 衍射强度数据的收集

要测定晶胞中原子的分布，必须收集晶体衍射的全部衍射点，按劳厄方程将各点进行指标化，并测量每个衍射点的强度。在应用旋进照相法收集强度数据时，一般采取沿某一晶胞轴的方向，如 c 方向，分层记录全部衍射花样，$hk0$，$hk1$，…，hkn。在每一层底片上衍射点的黑度高、面积大，则代表该衍射点的强度大。黑度的高低代表了积累衍射线的能量大小。为了收集一套高质量的强度数据，必须注意提高X射线管的效率，增加X射线的强度。用高灵敏度的感光胶片，尽可能降低胶片背景的黑度，仔细选择好晶体，调整好晶体的取向，并要特别注意各项实验操作，尽可能消除各种偶然误差。

测量衍射点的黑度，现在一般都用测微光度计。由测量得到的每个衍射点的相对强度，其数值一方面与实验条件有关，另一方面决定于强度标准的取值大小。所以需要经过各层之间、各晶体之间的强度的统一、校正以及进行一系列因子的修正，最后还原到结构振幅$|F_{hkl}|$值。

4.5 魏森堡照相法

魏森堡(Weinserberg)相机亦是一种移动晶体、移动底片的相机，不过它所摄出的衍射图

是歪曲的倒易格子层。它的主要用途是用来测定小分子的晶体结构，一般由于它的相筒半径较短，很难在生物大分子晶体结构分析中应用，但是，一种以同步辐射为光源的无屏大半径相筒的魏森堡相机，在收集生物大分子晶体衍射数据中已起到重要作用，所以掌握魏森堡相机摄取衍射图的基本原理，对理解 X 射线衍射法的基本原理和实验，仍然很重要。

魏森堡相机的基本装置见图4.17。装 X 光底片的相筒直径(内径)为57.3mm，可在水平方向移动；装晶体的转轴，可做360°的旋转。晶体绕轴每转2°，相筒位移1mm。层线屏直径(外径)为51mm，上有一条狭缝，只允许一层衍射圆锥通过。

摄取衍射图时，在装晶体的测角头上首先校准晶体的取向，然后利用层线屏分离不同层的衍射线，仅让一层衍射圆锥通过狭缝。当 X 射线入射到晶体上时，晶体慢慢地绕轴旋转，而层线屏外的圆筒形底片同步地沿着转轴方向移动，使在同一层上的倒易直线点阵有规律地展开在底片上。

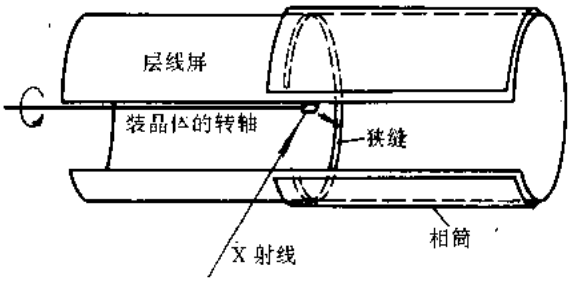


图4.17　魏森堡相机主要部件示意图

现在假定准备收集 c 轴方向的零层衍射花样，即晶体绕 c 轴转动，并且仅让 $hk0$ 衍射线通过层线屏狭缝。

图4.18中 $OQCF$ 表示0层倒易格子，其 OQ 倒易直线点阵通原点 O，处在起始位置，S 圆为反射球。当晶体转动 ω 角度后，OQ 直线与反射球相碰，P 点满足衍射条件，衍射产生，SP 方向

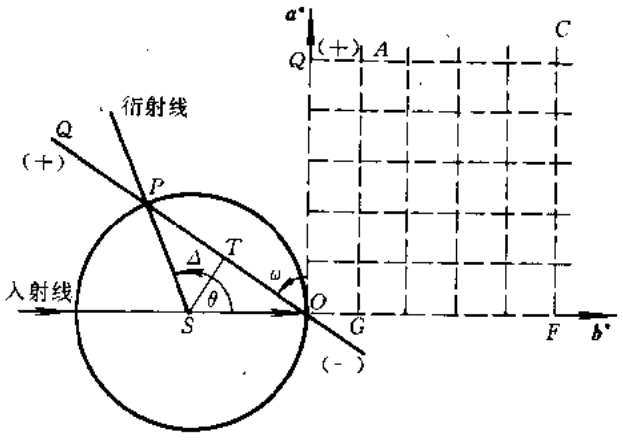


图4.18　通过原点的倒易直线点阵 OQ 与反射球相碰时的几何关系

即为衍射方向。这时入射线与衍射线交角为2θ，而$\theta=\omega$（因θ与$\angle TOS$，ω与$\angle TOS$互为余角），$\Delta=2\omega$。这个关系式说明，不管ω值多大，总是$\theta=\omega$，$\Delta=2\omega$。魏森堡相机就是按此关系设计的。

相机上圆筒形底片直径为57.3mm，即底片周长为180mm。当晶体绕轴转动90°时，相当于底片在垂直方向移动90mm，即晶体转动1°，底片在垂直方向移动1mm，而底片在水平方向位移1/2mm。如果把魏森堡衍射图的纵坐标定为Δ，横坐标定为ω，都用度数表示距离。由图4.18可以看到当OQ直线点阵转动$\omega<90°$时，直线扫过反射球的上半部；当$\omega=90°$时，OQ直线点阵与入射线平行；当$\omega>90°$时，OQ直线点阵扫过反射球的下半部。圆筒形底片在入射线进入处有一开口，当底片展开时，就出现斜直线OQ[图4.19(a)]。直线点阵OQ产生的衍射点位于一直线上，直线方程$\Delta=2\omega$，和中线交角63°26′，并在$\omega=90°$时有一转折；当$\omega=180°$时，OQ轴转到$-OQ$位置，即倒易直线点阵(+)轴与(−)轴互换了一个位置，而$-OQ$直线点阵的衍射轨迹与$\omega=0$时相同。

同理，通过原点O的OF直线点阵的衍射点的分布亦为一条斜直线，与OQ直线平行，相隔$\omega=90°$[图4.19(a)]。

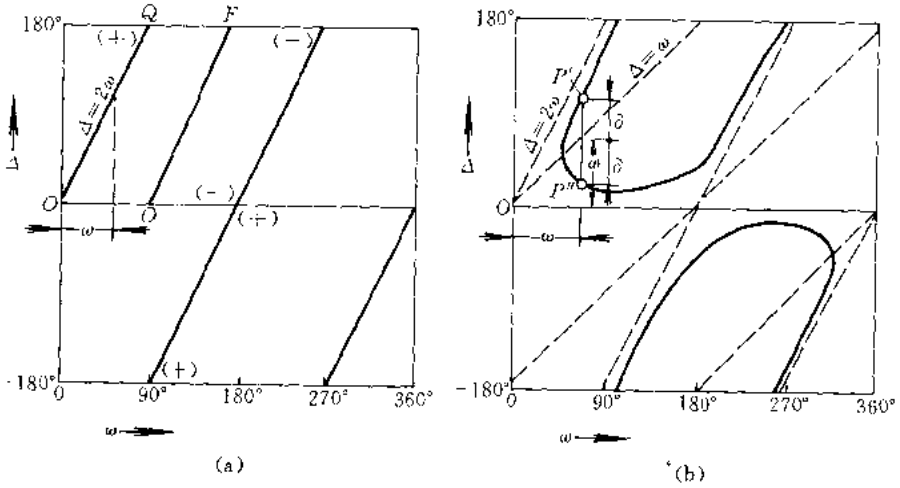


图 4.19

(a)通过原点的倒易直线点阵产生的衍射点的分布；

(b)不通过原点的倒易直线点阵产生的衍射点的分布

不通过原点O的倒易直线点阵GA，与OQ相距$d=OG'$。当转动ω角后，OQ线交于P点。而假定GA与反射球相交于P'和P''点(图4.20)，则由S所作的垂线ST，将平分$\angle P'SP''$为2个δ角，$\angle COP$为ω角，而$\angle ESO=\omega$，则

$$\angle OSP' = 2\theta' = \omega + \delta$$
$$\angle OSP'' = 2\theta'' = \omega - \delta$$

而ST直线方程$\Delta=\omega$。

GA倒易直线点阵的衍射点的轨迹是以直线$\Delta=\omega$为对称线的曲线。在同一ω值时，按$\Delta=\omega\pm\delta$而得两衍射点P'和P''[图4.19(b)]。

因此，整个0层倒易格子的衍射点都分布在如图4.21的网格交点上。当确定了倒易晶胞的 a^* 轴(OQ 轴)和 b^* 轴(OF 轴)后，测量两轴间的距离(度数)就可求出 γ^*。同理，亦可通过摄取另外方向的魏森堡图求得 a^*，β^*。晶轴的长度和夹角就可以通过倒易晶格与真实晶格之间的关系求得。

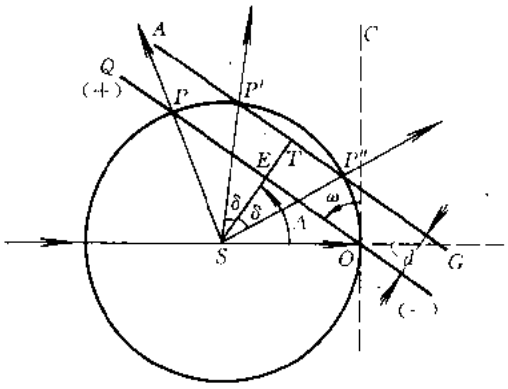


图4.20 不通过原点的倒易直线点阵 GA 与反射球相碰时的几何关系

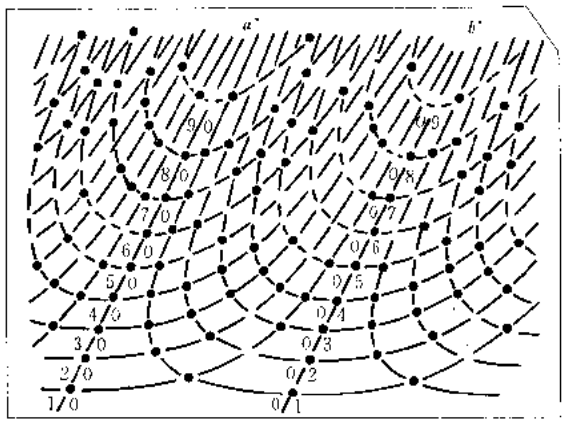


图4.21 $hk0$衍射点分布在网格上的示意图

高层魏森堡的衍射花样与零层的基本相似，仅因高层倒易格子离零层有一定距离，因而与反射球相截时，在图4.19中 δ 角将有变化，曲线的间隔也将略有改变[2][12]。

4.6 四圆衍射仪的基本原理

四圆衍射仪是由计算机控制的大型分析仪器。它使定向晶体、测量晶胞参数、测量“反射”强度等完全自动化，提高了衍射数据的质量和减轻了收集衍射强度数据的劳动。

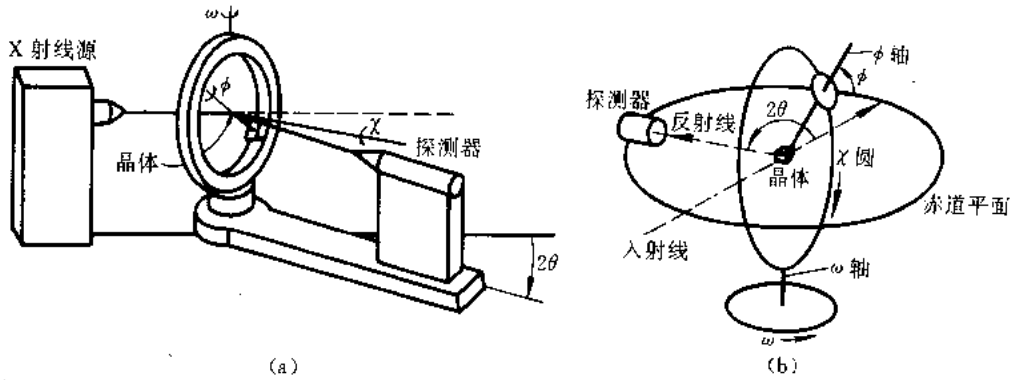


图4.22 四圆衍射仪
(a)仪器结构图；(b)四圆示意图

四圆衍射仪的结构原理示于图4.22。晶体定向由 ϕ，χ，ω 三个圆决定，检测器由第四个 θ

圆决定。入射 X 射线与 θ 圆同处于赤道平面内，ϕ 圆位于 χ 圆上。χ 圆垂直于赤道平面，由 ω 圆驱动，ω 圆与 χ 圆的轴彼此垂直。ω 角指示晶体绕 ω 轴的转动，ω 轴始终与入射的 X 射线垂直。

晶体装在测角头上，测角头固定在 ϕ 轴上，ϕ 圆可以沿着 χ 圆运动。ϕ，χ 圆作为整体可以围绕 ω 轴转动。计数器可在赤道平面上运动，入射线垂直于 ω 轴。在 $\omega=0$ 时 χ 圆平面垂于 X 射线；$\chi=0$ 时，ϕ 圆的轴平行于 χ 圆的垂直直径；$\phi=0$ 的位置是任意的，但亦可规定与某一倒易格子轴平行时为0°。

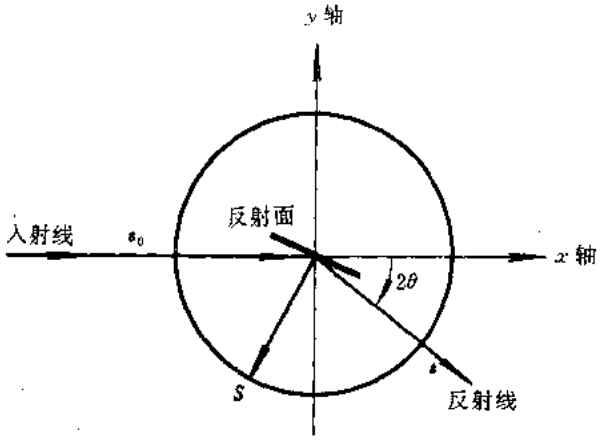


图4.23　示意仪器赤道平面和反射面组处在反射位置

晶体方向角 ϕ，χ，ω 和检测器的 θ 角的调整是按布拉格衍射条件决定的。当掠角（入射线与“反射”面的夹角）等于 θ_{hkl} 时，晶体的某一组平面就处在“反射”位置（图4.23）。根据布拉格公式，入射线 s_0，反射线 s 以及反射面的法线 S 必定处在同一平面内。在四圆衍射仪上，这个平面就是赤道平面。

在收集衍射强度数据时，要使每一个倒易格子点与反射球相碰，满足衍射条件，并在仪器的赤道平面上记录衍射强度，这就要求知道每个倒易格子点在倒易空间的坐标与仪器笛卡儿坐标的关系，然后再通过仪器坐标系与四个圆的角度关系，计算出每个衍射点的 ϕ，ω，χ，2θ 值，最后由机器使四个圆都到达正确位置，逐点收集强度数据。

衍射仪的笛卡儿坐标 x,y,z；x 轴的方向为入射线方向；z 轴垂直向上，与 ω 轴重合，y 轴的方向由右手定则确定。坐标系的原点处在入射 X 射线与四圆的 ω 轴在空间上的相交点上。

假定在仪器上已把晶体的方位调整，晶体重心已与衍射仪中心重合；四个圆都处于零点位置，这时任何一个倒易点的向量 $\boldsymbol{S}$ 为：

$$\boldsymbol{S}=ha^*+kb^*+lc^*=xi+yj+zk$$

i,j,k 为衍射仪笛卡儿坐标轴的单位向量。将 a^*,b^*,c^* 按 i,j,k 分解，得：

$$a^*=a_x^*i+a_y^*j+a_z^*k$$
$$b^*=b_x^*i+b_y^*j+b_z^*k$$
$$c^*=c_x^*i+c_y^*j+c_z^*k$$

代入上式得：

$$\boldsymbol{S}=(ha_x^*+kb_x^*+lc_x^*)i+(ha_y^*+kb_y^*+lc_y^*)j+(ha_z^*+kb_z^*+lc_z^*)k$$

$$= xi + yi + zk$$

由上可得：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} h \\ k \\ l \end{bmatrix}, \qquad R = \begin{bmatrix} a_x^* & b_x^* & c_x^* \\ a_y^* & b_y^* & c_y^* \\ a_z^* & b_z^* & c_z^* \end{bmatrix}$$

R 称为方位矩阵，它包含了晶体的方位和倒易晶胞信息。由已知晶胞参数，可求得方位矩阵，然后把任一衍射点 hkl，通过方位矩阵转换成衍射仪的坐标 (x,y,z)，再由仪器计算出四个角度值。

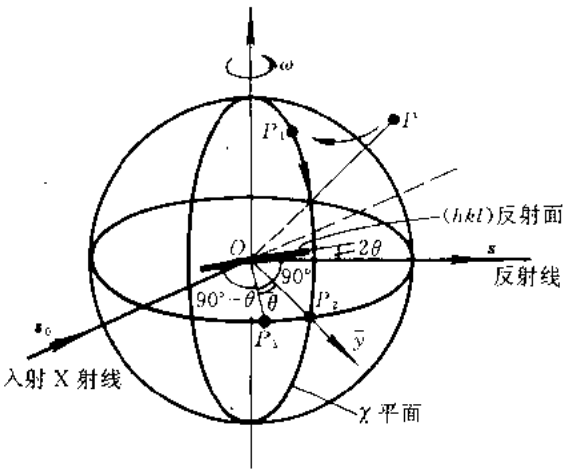


图4.24 示意某一倒易点进入反射位置的过程

假定在开始收集衍射强度时，四个圆都处在零位，当要使某一“反射”面组进入“反射”位置时，仪器按计算的该组面定向的四个角度值，首先驱动 ϕ 圆，转一个角度把该组面的倒易点 P 转到 ψ 平面的 P_1 处（图4.24），然后 ψ 圆转一个角度，把 P_1 点转到赤道平面上 P_2 处，接着 ω 圆转一个 θ 角，使 $\overrightarrow{OP}_2$ 向量随着 χ 圆平面转到 $\overrightarrow{OP}_3$ 位置。这时 $\overrightarrow{OP}_3$ 与“反射”面的法线重合，入射线与该“反射”面交角满足布拉格公式，衍射产生，检测器在 2θ 角位置记录“反射”线。

在四圆衍射仪上收集强度数据时，一般要经过安装、调整晶体，寻峰，测定倒易晶胞；计算衍射指标，修正方位矩阵；开始收集衍射强度数据等步骤[4][3]。

4.7 生物纤维状物质的 X 射线衍射

4.7.1 纤维衍射图

在生物大分子物质中有一类是天然纤维状的分子，如纤维蛋白、核酸等。其中纤维蛋白一般不溶于水，因此要得到具有三维排列的单晶几乎是不可能的。但是，如果把像羊毛、头发、蚕丝等天然物质直接进行回摆照相，就能得到 X 射线衍射图。这种图的分辨率很低，即衍射点都分布在入射线周围，不仅衍射点数量少，而且弥散、加宽，有的成短弧状（图4.25）。这种衍射图与单晶回摆图有明显的不同，是纤维状物质样品所特有的衍射花样，故称为纤维图。

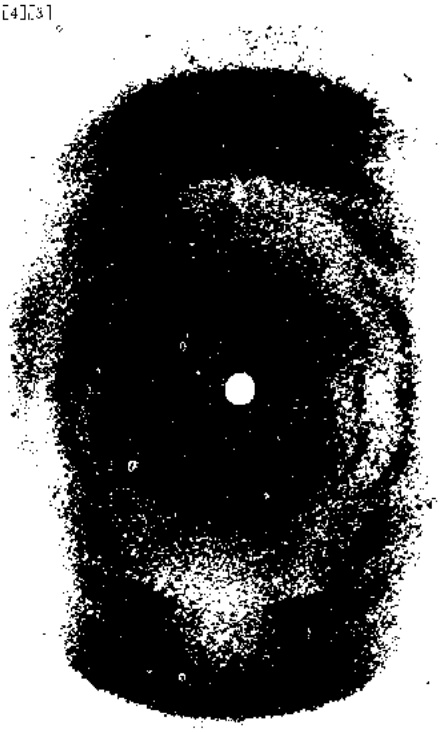


图4.25 DNA 纤维图的特点

纤维图与单晶回摆图的差别，反映了两种样品在结构上的不同。单晶是由一个空间格子所贯穿的、分子在三维空间有序排列的固体。这种三维有序结构可以划分成为数无限、不同间距、不同定向的平面组，而且，每一组晶面的重复周期是无限的。而在天然存在的上述纤维状物质中，纤维在某一区

域形成有序排列的晶态(小晶体),而在另一区域为无序的非晶态。因此整个样品是晶态、非晶态的混合体;有的样品在某一特定方向上小晶体排列是有序的,有的在二维排列有序,或者样品经过适当处理以后,小晶体在某一方向排列有序[图4.26(b)],但是在这种特定方向上晶面排列的有序性一般亦不是无限的。因此纤维状物质的衍射花样的特点决定于样品中晶态区的相对定向,即依小晶体沿某一特定方向排列的分布特征而定。下面按生物样品中可能出现的三种不同分布状态来讨论衍射图的特点。

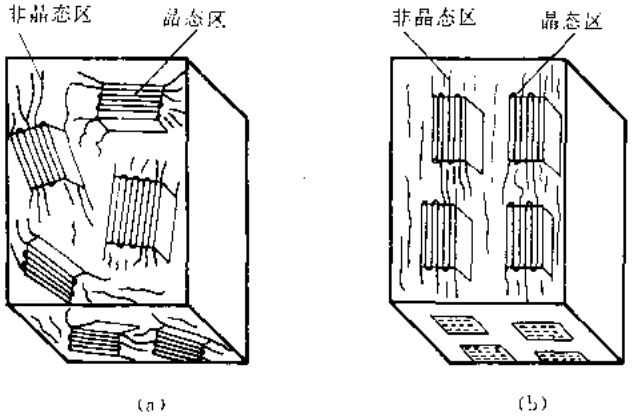


图4.26 纤维样品的特征

(a)示意纤维样品具有的晶态和非晶态区;(b)经适当处理后晶态区排列有序

第一,在样品中小晶体在 a,b,c 三个方向排列较有序,并近似无限,如图4.27(a)的分布状态。当把这样的样品进行X射线回摆照相时,就可得到类似单晶的衍射花样[图4.28(a)]。

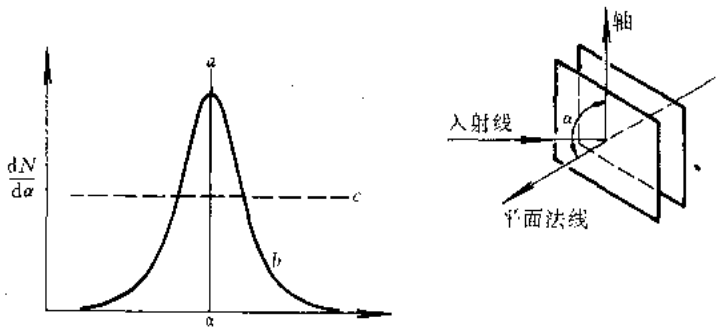


图4.27 样品内晶面定向的可能分布

α 角表示平面法线与摆轴间的夹角,$dN/d\alpha$ 表示在 α 至 $\alpha+d\alpha$ 之间晶面的数量;a 状态,晶面定向不随 α 角变化而变化,将得到类似单晶的衍射花样;b 状态,将得到短弧状的;c 状态,将得到同心圆的衍射花样

第二,如果在样品中的小晶体的数量足够多,而且晶体的定向是自由的、随机的分布,类似晶态粉末,如图4.27中 c 状态。当把这样的样品进行X射线回摆照相时,一定有许多小晶体处在满足布拉格公式的"反射"位置,而产生"反射"点。当然亦有很多小晶体处在不能满足"反射"条件的位置,但是由于有足够多数量的小晶体,而且随机定向,因此具有同一间距的平面组在产生"反射"时将具有相同的 θ 角,衍射点必定围绕入射线分布,假如有足够多数量的"反射"

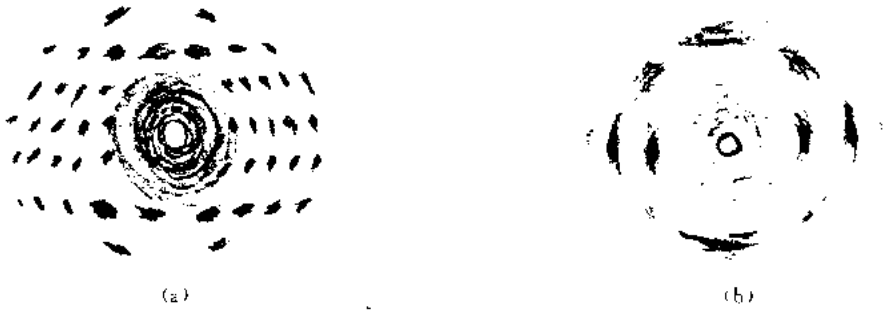


图4.28　不同纤维样品的衍射图
(a)高度有序纤维样品；(b)一般纤维样品

点，就可能形成一个围绕入射线的同心圆。同心圆中半径最小的，代表着具有最大晶面间距(d)的那些组"反射"面；半径最大的，代表着最小 d 的那些组面(图4.29)，并且不同级的"反射"与不同 d 的平面组所产生的"反射"线都混合在一起。这样的衍射图称为粉末衍射图。由粉末图可以根据布拉格公式中 $d_{(hkl)}$ 与 θ 角之间的关系求得晶体平面组的间距。

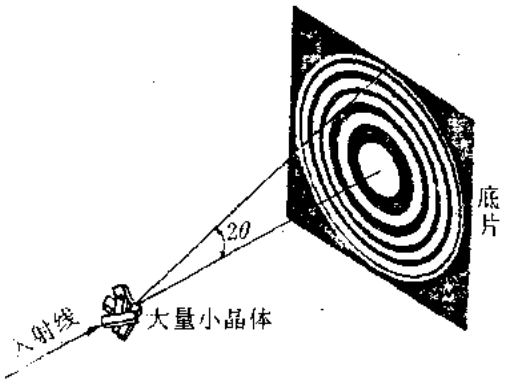


图4.29　由大量随机排列的小晶体产生的衍射花样(粉末图)

第三，在大多数生物纤维状物质样品中晶态区的排列处在上述两种状态之间，样品中的绝大多数小晶体的取向仅在某一方向有一个较大的如图4.27中(b)状态的概率分布。因此在拍摄这种样品的回摆图时，仅在一定的 θ 角度范围内小晶体能满足布拉格公式而产生"反射"，同时由于绝大多数小晶体围绕纤维轴定向，因而不可能形成像粉末图那样的同心圆衍射花样，而仅能形成短弧状的衍射斑点，而且由于小晶体间镶嵌性很差，以及混有非晶态区，因而衍射斑点弥散。衍射点的数量亦随 θ 角的增大而迅速减少，整个衍射图层次不清，形成生物纤维状物质特有的衍射花样——纤维衍射图(图4.28(b))。

4.7.2　螺旋结构的纤维衍射图

在生物大分子的纤维结构中常具有螺旋结构，每个类似的亚基围绕着螺旋轴排列。假如在平行于 a 方向上每个纤维状分子具有一个六重螺旋轴的结构，相邻原子团或亚基的平移距为 $a/6$，而螺旋的螺距为 a (图4.30左)。根据3.2.6节讨论的系统消光规律，在这种样品的衍射图上，100，200，300，400和500衍射将由于相邻原子团散射线间的光程差不为波长的整数倍而相互干涉，结果衍射点的强度为零，仅600衍射，以及 h 为6的倍数的各衍射点强度不为零(图4.30右)。但是在非 $h00$ 方向上，如 $1kl$，$2kl$，…，hkl 等衍射中仍将有衍射点出现，在图4.31(a)中纤维分子在 a 方向重复周期为 a，在 b 方向上假定纤维分子排列有序，重复周期为 b，并构成平面晶胞为 $Ob'a'O'$。在这样的螺旋结构的晶体格子中，可能存在的平面组为(010)，($1\bar{1}0$)，

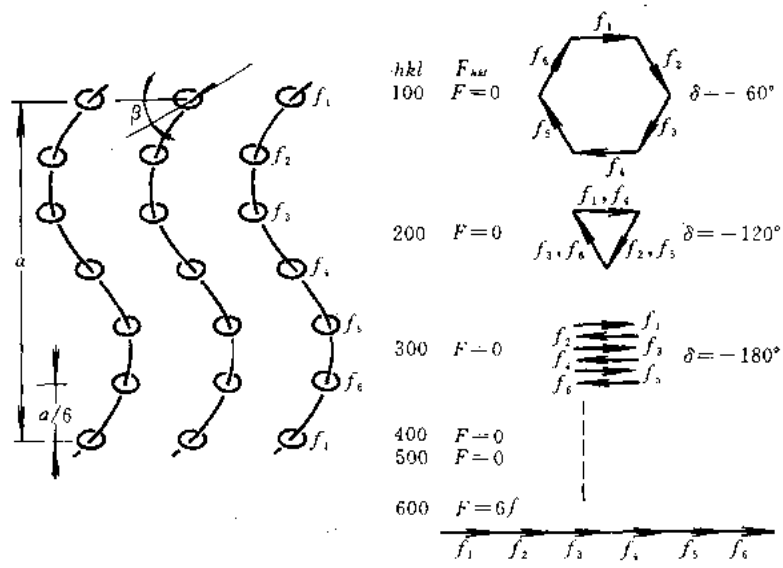


图4.30 具有六重螺旋轴的结构和消光规律

图左为螺旋结构，右为结构因子向量图

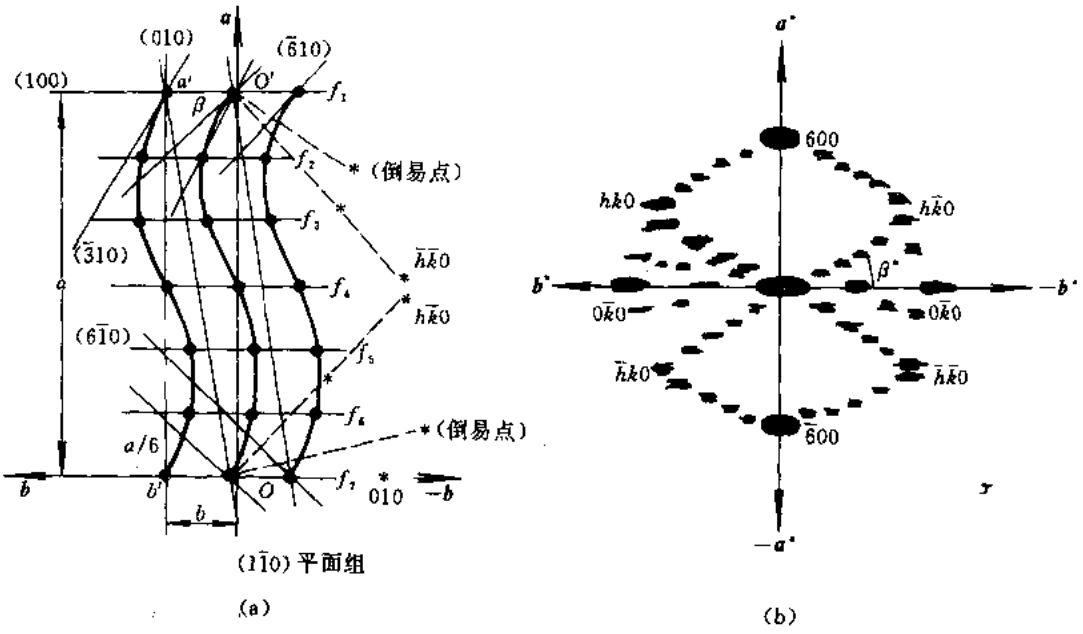


图4.31 螺旋结构的衍射花样

(a)各平面组的倒易点的分布；(b)螺旋结构衍射花样的示意图

(3̄10)，(6̄10)…等等。由于实际的纤维样品不是单晶，因此可能存在的晶面组数量以及每一平面组内的平面重复周期数都较少。当上述样品进行X射线实验时，各平面组都将可能产生各自的“反射”，例如与 a 轴相平行的(010)平面组产生的各级“反射”(0k0型衍射)将分布在水平位置上(b^* 轴上)，而与 a 轴或 b 轴相交成一定角度的平面组所产生的“反射”(hk0型)将分布

在 a^* 与 b^* 轴之间。而在上述平面组中指标最高的$(6\bar{1}0)$平面组与 a 轴所交的角度最大，它产生的各级"反射"将分布在靠近 a 轴的一侧[图4.31(a)]并垂直于$(6\bar{1}0)$平面组。同时由于晶面指标高的晶面组内的晶面间距小于指标低的晶面间距，因此高指标的晶面组的衍射点分布在离晶格原点远的地方，而低指标的晶面组的衍射点将分布在离原点近的地方。根据衍射花样中心对称定律，在上述衍射图中还将出现$\bar{h}\bar{k}0$型衍射点。同理，晶体格子中可能存在的$(\bar{6}\bar{1}0)$，$(\bar{3}\bar{1}0)$…等平面组亦将可能产生相应的"反射"($\bar{h}\bar{k}0$和 $hk0$型)。因此整个螺旋结构样品的衍射花样将在 a^* 轴方向和与 b^* 轴成 β^* 角(螺旋角)的区域内无衍射点[图4.31(b)]。这种衍射花样的特点是一切具有螺旋结构纤维的 X 射线回摆图的共同特征。

4.7.3 纤维状物质的结构分析

如果通过照相法拍摄了分辨率较高的纤维衍射图，就可对衍射花样上衍射点的分布(点的几何位置、对称性等)进行分析和指标化，并进一步计算出在某一方向上纤维结构的重复周期，以至推测出晶胞的大小和形状。通过测量衍射点的强度以及对强度进行适当处理后就可得到一定数量(一般为几十至上百个)的衍射强度实验值(I_o)。在此基础上就可试用模型法(试差法)来分析纤维物质的分子及晶体的结构。

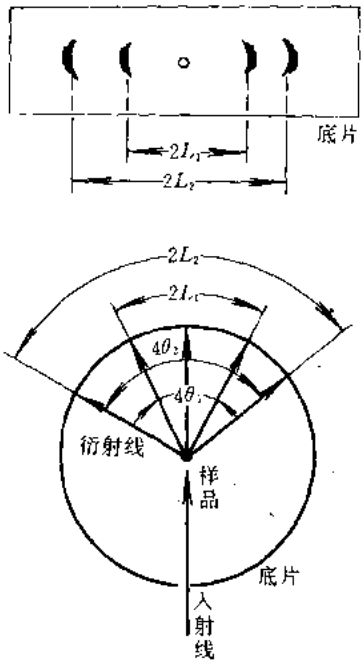


图4.32 短弧间距$2L$与θ角之关系

4.7.3.1 重复周期的测量

在展平的圆筒形底片上衍射花样中的每对近似于短弧的衍射点，如图4.32所示。在图中每对短弧间的距离$2L$与产生这些短弧线的"反射"面组之θ角之间，存在一定的关系，例如在水平位置上的一对短弧形衍射点与产生该对衍射点的"反射"面组之θ角间的关系如下：

$$\frac{4\theta}{2L}=\frac{360^\circ}{2\pi R},\text{即}\ \theta=\frac{L}{\pi R}90^\circ$$

式中 R 为圆筒形底片之半径。因为底片的圆周 $2\pi R$ 所对的圆心角为360°，而 $2L$ 所对之圆心角应等于相应“反射”圆锥面(“反射”线)之张角 4θ [图4.32(下)]。上式中 R 是相筒半径，为已知值，而每对“反射”弧之间距 $2L$ 可从衍射图上测量，因此可由上式求出相应“反射”面组之 θ 角值，然后应用布拉格公式 $2d\sin\theta=n\lambda$，即可求出“反射”面组中平面间距(重复周期) d/n 。

4.7.3.2 结构的测定

由于前述的纤维状物质样品本身的特点，因此所收集到的衍射点一般数量很少，质量亦较差，因而它的分析方法与一般单晶结构分析方法有很大差别，目前主要采用模型法，其大体步骤如下：制备纤维样品，然后进行 X 射线衍射照相，并对每个衍射点指标和测量强度(I_o)，同时提出分子模型和晶体结构模型，而模型的建立主要是根据化学、生物化学和晶体结构的一般规律，如多肽链中肽键的平面结构；多肽链在形成二级结构时，应保证形成最多的、可能的氢键；以及在形成晶体时，分子堆积应遵循最密堆积、对称性最高、能量最低原则等等。然后根据提出的模型进行 X 射线衍射的理论强度 I_c 的计算，并与实验值 I_o 相比较。按比较结果修正初始模型，接着再计算修正模型的 I_c，再与 I_o 比较和修正模型。经多次、反复进行上述过程，直至 I_o 与 I_o 最好地一致为止。这时的分子和晶体结构模型即为待测纤维物质的分子和晶体结构模型。这里必须强调在实际结构分析工作中，其过程和重复的次数要复杂和多得多。[6][6][11][18]

第五章　晶胞中原子位置的测定

晶胞中原子的分布决定衍射花样的强度分布，每个原子的坐标信息存在于晶体的衍射强度之中，而要通过衍射强度数据来测定晶胞中每个原子的坐标，则要通过电子密度函数的计算和解释。

5.1　晶体的电子密度分布函数

晶体结构可理解为一定种类和数目的原子按一定周期进行重复排列的、分立的分布，但也可表达为具有一定重复周期的电子密度的连续分布。在原子中心的电子云密度最高，无原子的地方电子云密度为零。整个晶胞内原子的分布可以看成是由不同密度的电子云分布。假如把整个晶胞内的电子云密度分布计算出来，而且每个电子云密度的峰可以辨认，则晶胞中的每个原子的分布亦就测定了。不同种类的原子，原子序数不同，电子数不同，核外电子数多，原子序数高，则峰值亦高。峰在晶胞中的精确位置即为原子的坐标。

任何一个周期函数都可以用正弦、余弦函数来表达，电子密度分布函数是一个周期函数，同样可以这样来表达。

设有一个一维周期函数 $f(x)$，周期为 a，则

$$f(x) = a_0 + a_1\cos 2\pi x + a_2\cos 2\pi(2x) + \cdots + a_n\cos 2\pi(nx) + b_1\sin 2\pi x + b_2\sin 2\pi(2x) + \cdots + b_n\sin 2\pi(nx)$$

$$f(x) = a_0 + \sum_1^n (a_h\cos 2\pi hx + b_h\sin 2\pi hx)$$

又

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}$$

$$\sin x = \frac{-i(e^{ix} - e^{-ix})}{2}$$

代入上式得

$$f(x) = a_0 + (1/2)\left[a_1e^{2\pi ix} + a_1e^{-2\pi ix} + a_2e^{2\pi i(2x)} + a_2e^{-2\pi i(2x)} + \cdots\right] - (i/2)\left[b_1e^{2\pi ix} - b_1e^{-2\pi ix} + b_2e^{2\pi i(2x)} - b_2e^{-2\pi i(2x)} + \cdots\right]$$

$$f(x) = a_0 + (1/2)\left[(a_1 - ib_1)e^{2\pi ix} + (a_2 - ib_2)e^{2\pi i(2x)} + \cdots\right] + (a_1 + ib_1)e^{-2\pi ix} + (a_2 + ib_2)e^{-2\pi i(2x)} + \cdots$$

$$f(x) = \sum_{-n}^{n} C_he^{2\pi ihx}$$

在这里 $C_h = (a_h - ib_h)/2$，$C_{\bar{h}} = (a_h + ib_h)/2$，$C_0 = a_0$，这是 $f(x)$ 函数的指数表达形式。或者

$$f(x) = \sum_{-n}^{n} C_h(\cos 2\pi hx + i\sin 2\pi hx)$$

由此一个一维电子密度函数亦可表达为

$$\rho(x) = \sum_{h'} C_{h'} e^{2\pi i h' x} \tag{1}$$

在此 h' 是从$-\infty$至∞的整数。

而结构因子公式

$$F_h = \sum_{j=1}^{n} f_j e^{i2\pi h x}$$

如果用连续的电子密度函数 $\rho(x)$ 代替每个分立的原子，则

$$F_h = \int_0^1 \rho(x) e^{2\pi i(hx)} a\,dx$$

a 为周期。把(1)式代入上式，则

$$F_h = \int_0^1 \left(\sum_{h'} c_{h'} e^{2\pi i h' x} \right) e^{2\pi i h x} a\,dx$$

$$= a \sum_{h'} c_{h'} \int_0^1 e^{2\pi i(h+h')x} dx$$

当 $h' = -h$ 时

$$\int_0^1 e^{2\pi i(h+h')x} dx = 1$$

而当 $h' \neq -h$ 时

$$\int_0^1 e^{2\pi i(h+h')} dx = 0$$

因此

$$F_h = aC_{\bar{h}}$$

$$C_{\bar{h}} = F_h / a \tag{2}$$

把(2)式代入(1)式，并以$-h$代取 h'，则

$$\rho(x) = \frac{1}{a} \sum_h F_h e^{-2\pi i(hx)}$$

这是电子密度函数在一维情况下的表达式，推广到三维空间的电子密度函数表达式为

$$\rho(xyz) = \frac{1}{V_0} \sum h \sum_{-\infty}^{+\infty} k \sum l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

式中 V_0 为晶胞体积，hkl 为衍射指标，x,y,z 是晶胞中某处的坐标，而 $\rho(xyz)$ 为距离晶胞原点 x,y,z 处的电子密度。F_{hkl}为结构因子，由 $F_{hkl} = |F_{hkl}| e^{i\varphi_{hkl}}$ 关系式可以看出 $|F_{hkl}|$ 能从衍射强度数据中推引出来，而 φ_{hkl} 却无法从实验中得到。因此无法从衍射强度数据直接算出结构因子 F_{hkl}，也就无法直接计算电子密度函数，因而无法得知原子在晶胞中的坐标。这就是晶体结构分析中的相角问题，有待以后讨论。

5.2 电子密度图

假如已经收集了晶体的衍射强度数据，并且经过修正、统一和还原为结构振幅 $|F_{hkl}|$ 的数值，以及通过间接的方法测定了每一个衍射的相角 φ_{hkl}，这时就可按电子密度函数公式计算晶胞中每一点的电子密度。在实际工作中可计算三维电子密度分布，亦可计算某一方向的截面或

投影。

在计算三维电子密度时可按下列公式进行

$$\rho(xyz)=\frac{1}{V_0}\sum h\sum k\sum l F_{hkl}e^{-2\pi i(hx+ky+lz)}$$

或者按其他形式的公式计算，如用 $F_{hkl}=|F_{hkl}|e^{i\alpha_{hkl}}$，$F_{hkl}=|F_{hkl}|\cos\alpha_{hkl}+i|F_{hkl}|\sin\alpha_{hkl}$或者 $F_{hkl}=A_{hkl}+iB_{hkl}$(此处 $A_{hkl}=|F_{hkl}|\cos\alpha_{hkl}$，$B_{hkl}=|F_{hkl}|\sin\alpha_{hkl}$)作为电子密度函数的系数，即以等式的右项代替 F_{hkl}，并假定衍射强度分布符合中心对称定律(*Friedel* 定律)，以及经归并、整理后可得：

$$\rho(xyz)=\frac{1}{V_0}\sum h\sum k\sum l|F_{hkl}|\cos 2\pi[(hx+ky+lz)-\varphi_{hkl}]$$

这一形式使得每一个相角 φ_{hkl} 与每一 $|F_{hkl}|$ 的关系更清晰。

在计算电子密度函数的截面时，通常将 x,y,z 中任意一个固定为某一常数，如 $z_1=0$，$z_2=\frac{1}{2}$ …等任意一个数值，然后计算二维截面，这样便于绘图和表示出三维电子密度的分布。例如计算 $z_1=0$ 的截面，按公式：

$$\rho(xyz)=\frac{1}{V_0}\sum h\sum k\sum l F_{hkl}e^{-2\pi i(hx+ky+lz)}$$

$$=\frac{1}{V_0}\sum h\sum k[\sum l F_{hkl}e^{-2\pi ilz}]e^{-2\pi i(hx+ky)}$$

当 $z_1=0$ 时，公式简化为

$$\rho(xy0)=\frac{1}{V_0}\sum h\sum k[\sum l F_{hkl}]e^{-2\pi i(hx+ky)}$$

在计算晶胞中每个点 (xyz) 的电子密度时，一般是按照一定间隔 $(\Delta x,\Delta y,\Delta z)$ 逐点按公式计算。因此计算机算出来的每一层电子密度是一套数值，每个数值代表在晶胞中某点的电子密度。电子密度极大值的位置，即是原子的中心位置。原子的电子数越多，电子的密度也越大。据此，原则上可从电子密度分布中来分析确定各种原子的位置。

为了突出原子的形状及整个结构，可在电子密度的平面图上把电子密度数值相同的点，例如 $0.04e/nm^3$ 用线连接起来，绘成等高线图，称为电子密度图。图 5.1 示意了在(28/60,19/60)处有一个电子密度峰。假如能够辨认出该峰为某一原子，就可由图上读出该原子在晶胞中的坐标。

X 射线衍射分析的分辨率越高，电子密度图上反映出的晶体结构的细节越清楚。在这里要说明一下，晶体结构分析中的分辨率，与一般了解的光学显微镜的分辨率有所不同，这里的分辨率是指在晶体收集衍射强度数据时，所记录的衍射点反映的最小平面间距 d_m 的大小。按公式 $2d\sin\theta=n\lambda$，得 $d=\frac{n\lambda}{2\sin\theta}$，因此当 $n=1$，λ 为0.154nm(CuK_α)时，假如 θ 角接近90°，为理论上的最小分辨率间距 $d_m=0.075$nm。而在实际收集蛋白质晶体的衍射强度数据时，一般 θ 值仅为30°左右，因而实际所能达到的分辨率要低得多，而在电子密度图中能达到的分辨率则更低。这是由于生物大分子晶体内部分子排列的有序性较差，以及衍射数据和相角计算的质量等因素的影响。

在蛋白质晶体结构分析中，一般把 d_m 大于0.45nm 的称为低分辨率的电子密度图，d_m 小于0.35nm、大于0.25nm 的称为中等分辨率的电子密度图，d_m 小于0.2nm 的称为高分辨率的

电子密度图。

在低分辨率的电子密度图中，肽链的α螺旋区可以看清，而肽链的其他部分大体上能分辨，但是较零散，因此在低分辨率的电子密度图中整个肽链走向不能被追踪。金属原子，如胰岛素中的锌原子，肌红蛋白中的铁原子具有高的电子密度和较大体积，很容易分辨(图5.2)。

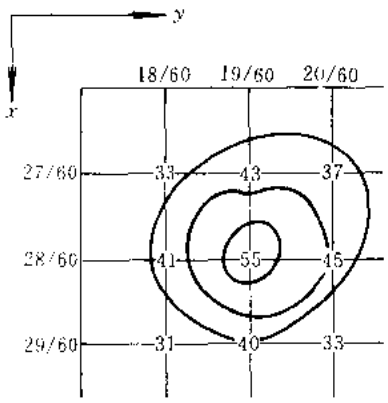


图5.1 示意部分电子密度图中不同位置的电子密度值

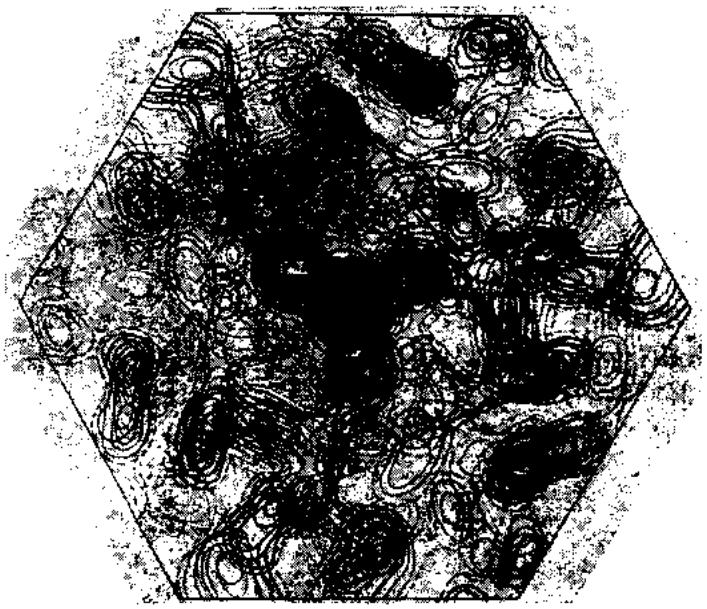


图5.2 在10- 18/60胰岛素电子密度截面叠合图上表现出来的锌离子的配位情况(锌在图中心)

在中等分辨率的电子密度图中整个多肽链的走向可以辨认，但是具体地分辨每个氨基酸残基较困难。在0.3—0.35nm的分辨率图中，肽链的羰基，在比较有序的部位可以分辨清楚，但在肽链最柔软的部分，即使在0.25nm分辨率的图上亦不可能看清楚。在0.3nm分辨率的电子密度图中，一般肽链的侧链能够看清，而在分子表面的侧链基团由于基团定向较差，而分辨不清，如赖氨酸残基的侧链仅β碳原子能看清。苯丙氨酸、酪氨酸、组氨酸、异亮氨酸、极性的天门冬氨酸、天门冬酰胺、谷氨酸、谷氨酰胺、苏氨酸等侧链要在0.28nm分辨率时，才能看清楚，二硫桥在0.28nm时能看清，但是硫原子还不能分辨清楚。

在小于0.2nm的高分辨率电子密度图中，如在0.19nm的胰岛素晶体结构的电子密度图中，二硫键的两个硫原子能分辨，β碳原子可以较精确的定位；芳香族的残基更好确定，环的中心电子密度明显降低，环能看清，α侧链更清楚地与主链分开。在较好的电子密度图中，肽链的构象能够很好地确定。图5.3表示了在不同分辨率下，同一氨基酸残基的电子密度图。

5.3 电子密度图的解释与原子在晶胞中的位置

由于蛋白质分子量很高，原子数量很多，以及蛋白质晶体内部分子排列的有序性较差等原因，要得到中等和高分辨率的电子密度图是相当困难的，而电子密度图的解释亦是相当复杂的工作，是一个由粗到细，由易到难，综合分析，逐步精化的过程。一套中等或高分辨率的电子密

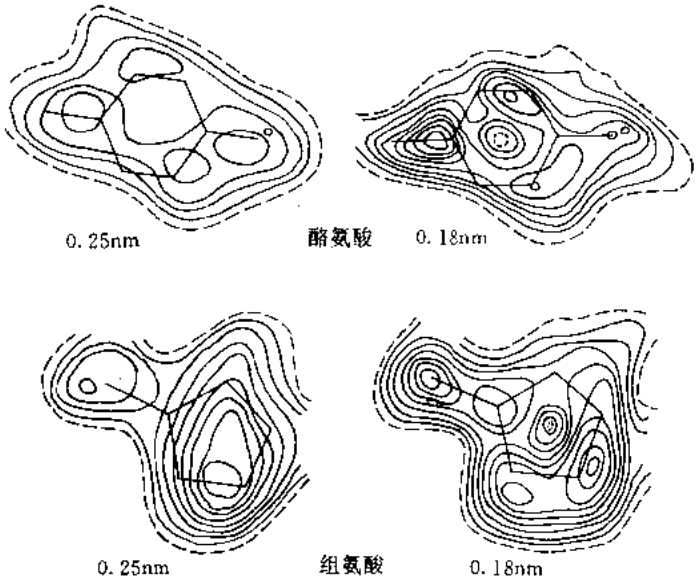


图5.3　在不同分辨率的电子密度图上，酪氨酸、组氨酸的电子密度图

度图的解释工作，一般要经过几个步骤：首先，在透明塑料片上绘出按一定比例的各层电子密度图，如按1cm＝0.3nm 的比例。第二，按共同原点和准确的间距把各层电子密度图叠合起来，构成三维电子密度图(图5.4)，初步观察分子的边界、非结晶学对称性以及肽链走向等总体情况，纠正一些明显的、不合理的偏差。第三，如得到较理想的结果，可再进一步缩小比例，如按2cm＝0.3nm 的比例绘制各层电子密度图，再叠合成三维电子密度图，在该图上，可根据分子的一级结构序列，直接与电子密度图作对照分析，观察肽链的走向，辨认像 α 螺旋部分，有明显标记的部分如金属原子结合处，酶与底物结合部位等，并逐步扩展，进而确定分子的部分结构模型。第四，在上述解释的基础上，再利用三维图像显示仪对电子密度图做进一步地解释和对结构模型做进一步修正。

三维图像显示仪是由电子计算机控制的、可运行多种程序，具有高分辨率屏幕的显示仪器，它可提供构建、显示、修改和操作二维和三维的高质量的、层次分明的彩色图像，例如根据操作者的要求，可按氨基酸或肽键的标准键长、键角以及肽链的扭角等数据，构建各种氨基酸、肽链和整个蛋白质分子的模型，并可任意删除、增加或置换肽链中的一个或多个残基；可按要求改变某一肽段的构像；可按任一方向旋转或平移整个分子；可叠合两个分子或肽段进行比较研究；可把部分或全部三维电子密度图在图像显示仪上重演，以及与构建的分子模型重叠、比较(图5.5)等等。

因此在三维图像显示仪上，经过分子模型与电子密度图之间细致、精确的对照分析和调整，可确定部分或大部分乃至整个分子的结构模型，并由三维图像仪计算出部分或大部分、乃至整个分子的非氢原子坐标。但是由于在强度测量、相角的计算、同晶型质量等方面可能会出现的系统的或偶然的误差，以及分子模型还可能存在某些问题，仍然与电子密度图不完全相吻合，而影响结果的精度，因此由电子密度图得到分子模型以后，还需对原子坐标、温度因子、占有率和比例因子等采取电子密度函数法、最小二乘等方法进行修正，从而提高原子坐标的精

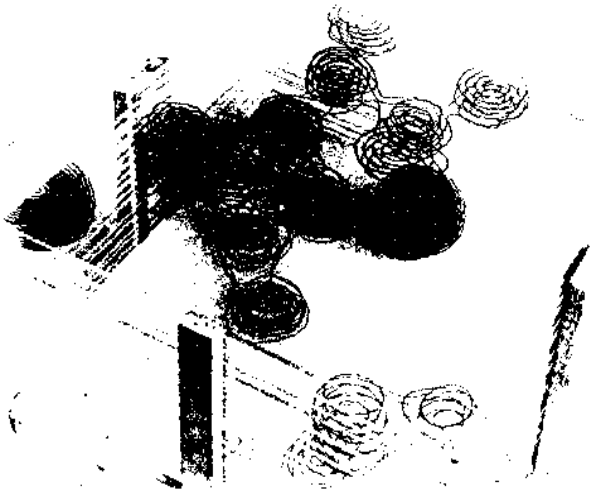


图5.4　苄青霉素钾电子密度截面叠合图

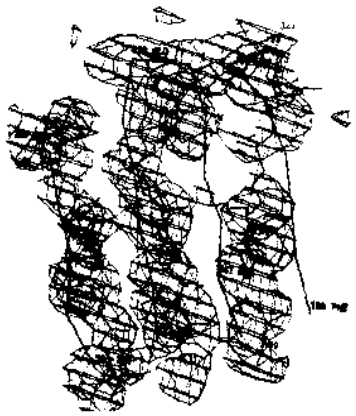


图5.5　在图像仪上显示去六肽卵清蛋白残基Glu342附近电子密度(细线)与α-1-蛋白水解酶抑制剂相应位置肽段(粗线)间的比较

度、改善相角的质量等。由分子模型的原子坐标可计算出每个衍射点的相角，应用此相角值和实验得到的结构振幅值(F_o)可再计算出电子密度图，并在图像仪上对图作进一步解释，与分子模型作对比、调整模型中的不够正确部分，然后对原子坐标等参数再修正，再计算电子密度图，再解释，再修正分子模型。如此反复，直至建立起与衍射强度数据完全吻合的分子模型，从而测定出全部原子的坐标(表5.1)。由原子坐标可计算出分子的键长、键角和扭角等结构数据。[2][7][10]

在晶体结构的分析过程中，从提出初步的分子结构模型到最后解出结构，需要进行各种修正，其修正正确性的判断主要以实验观测的结构因子振幅与理论计算的结构因子振幅相吻合为标准，即以偏差因子 R 来判断。

$$R=\frac{\sum_h |F_o(h)-F_c(h)|}{\sum_h F_o(h)}$$

F_o 为观察结构振幅；F_c 为计算结构振幅；h 代表衍射指标 hkl 。

在应用同晶置换法测定相角过程中，对重原子参数和相角修正时，R 因子为0.10左右，而在小分子晶体结构分析中，R 因子在0.10以下，好的结果可以达到0.05以下，这时原子坐标的精确度可达到0.0004nm，键长的精确度可到0.0006nm。

表5.1 猪胰岛素A链4个氨基酸残基的非氢原子坐标

原子编号	原子名称	氨基酸残基名称	氨基酸残基序列号	原子坐标 x (nm)	原子坐标 y (nm)	原子坐标 z (nm)	占有率	温度因子 $(nm)^2$
1	N	GLY A	1	−0.881	1.702	1.423	1.00	2.11
2	CA	GLY A	1	−0.989	1.700	1.324	1.00	1.92
3	C	GLY A	1	−0.966	1.558	1.263	1.00	3.94
4	O	GLY A	1	−0.973	1.461	1.334	1.00	2.85
5	N	ILE A	2	−1.028	1.555	1.134	1.00	2.75
6	CA	ILE A	2	−1.043	1.432	1.059	1.00	1.98
7	C	ILE A	2	−0.931	1.329	1.070	1.00	1.30
8	O	ILE A	2	−0.953	1.211	1.099	1.00	1.85
9	CB	ILE A	2	−1.090	1.448	0.914	1.00	5.55
10	CG1	ILE A	2	−1.153	1.314	0.868	1.00	3.70
11	CG2	ILE A	2	−0.976	1.486	0.819	1.00	2.75
12	CD1	ILE A	2	−1.283	1.291	0.948	1.00	2.49
13	N	VAL A	3	−0.809	1.376	1.045	1.00	1.62
14	CA	VAL A	3	−0.695	1.283	1.500	1.00	1.43
15	C	VAL A	3	−0.682	1.217	1.187	1.00	2.53
16	O	VAL A	3	−0.657	1.097	1.200	1.00	2.62
17	CB	VAL A	3	−0.570	1.367	1.015	1.00	1.79
18	CG1	VAL A	3	−0.440	1.288	1.032	1.00	3.08
19	CG2	VAL A	3	−0.586	1.418	0.871	1.00	2.72
20	N	GLU A	4	−0.700	1.300	1.290	1.00	1.66
21	CA	GLU A	4	−0.698	1.249	1.427	1.00	1.79
22	C	GLU A	4	−0.801	1.153	1.466	1.00	2.14
23	O	GLU A	4	−0.781	1.044	1.519	1.00	2.00
24	CB	GLU A	4	−0.688	1.371	1.522	1.00	1.90
25	CG	GLU A	4	−0.561	1.455	1.501	1.00	2.23
26	CD	GLU A	4	−0.567	1.542	1.374	1.00	1.62
27	OE1	GLU A	4	−0.678	1.594	1.345	1.00	2.05
28	OE2	GLU A	4	−0.458	1.554	1.311	1.00	2.46

第六章　解决相角问题的方法

解决生物大分子晶体结构分析中的相角问题：主要是应用重原子同晶置换法。这种方法是把重原子引进到由轻原子组成的结构中，以重原子作为标识原子来获得位相信息；如果结构中已存在较重的原子，可使用反常散射法；如果已有一个类似的生物大分子的晶体结构已经解出，就可用分子置换法；如果待测分子不但与已知晶体结构的分子在三维结构上类似，而且两者是同晶型的，还可以用差值帕特逊方法或差值傅利叶方法来得到位相信息；帕特逊函数法主要用于解决小分子的相角问题，但是它的基本原理在分子置换法、测定重原子在生物大分子晶体中的位置等方面仍具有很重要的作用；另外也有人在研究用直接法解决大分子的相角问题，这种方法是应用衍射点之间的数学关系来获得位相信息。

6.1　帕特逊(Patterson)函数法

帕特逊函数是用实验所得的衍射强度数据，计算出结构因子的振幅值，再以振幅的平方值 $|F_{hkl}|^2$ 为傅里叶级数的系数的函数。由这样的函数计算出的结果是结构的向量图，分析向量图可确定原子的坐标，由坐标可计算出原子散射波在某一衍射中的位相。在一维情况下，帕特逊函数的具体形式为：

$$P(u)=\frac{1}{a}\sum h\ F_h|^2 e^{-i2\pi(hu)} \tag{1}$$

a 为 $\boldsymbol{a}$ 方向原子重复周期，u 为帕特逊图上峰的分数坐标。上式可由帕特逊函数的定义推出来。

$$P(u)=\int_0^1 \rho(x)\cdot\rho(x+u)\cdot a\mathrm{d}x \tag{2}$$

式中：$\rho(x)$ 表示点 x 处的电子密度，即

$$\rho(x)=\frac{1}{a}\sum_{+\infty}^{-\infty} hF_h e^{-i2\pi hx}$$

$\rho(x+u)$ 表示点 $(x+u)$ 处的电子密度，即

$$\rho(x+u)=\frac{1}{a}\sum_{+\infty}^{-\infty} hF_h e^{-i2\pi h(x+u)}$$

代入(2)式得：

$$\begin{aligned}P(u)&=\int_0^1\left[\frac{1}{a}\sum_{+\infty}^{-\infty}h'F_{h'}e^{-i2\pi h'x}\right]\cdot\left[\frac{1}{a}\sum_{+\infty}^{-\infty}hF_h e^{-i2\pi h(x+u)}\right]\cdot a\mathrm{d}x\\&=\frac{1}{a}\sum_{-\infty}^{-\infty}h'\sum hF_{h'}F_h\int_0^1[e^{-i2\pi h'x}]\cdot[e^{-i2\pi hx}\cdot e^{-i2\pi hu}]\mathrm{d}x\\&=\frac{1}{a}\sum_{-\infty}^{-\infty}h'\sum hF_{h'}F_h e^{-i2\pi hu}\left[\int_0^1 e^{-i2\pi(h'-h)x}\mathrm{d}x\right]\end{aligned}$$

式中 h 为衍射指标，在括号中的 h 值有时相等，但在大多数情况下是不等的，故暂时加上“′”以

示区别。

式中积分，当 $h' = -h$ 时

$$\int_0^1 \mathrm{d}x = 1$$

当 $h' \neq -h$ 时，积分为0。将此积分结果代入上式得：

$$P(u) = \frac{1}{a}\sum_{-\infty}^{-\infty} h(F_h F_{\bar{h}})\mathrm{e}^{-i2\pi hu}$$

$$= \frac{1}{a}\sum_{-\infty}^{-\infty} h|F_h|^2\mathrm{e}^{-i2\pi hu}$$

推广到三维的情况

$$P(uvw) = \frac{1}{V}\sum h \sum_{-\infty}^{-\infty} k \sum l |F_{hkl}|^2 \mathrm{e}^{-i2\pi(hu+kv+lw)}$$

式中 V 为晶胞的体积，u,v,w 为帕特逊空间某一点的坐标。

由上面帕特函数的定义可以理解，在帕特逊空间中的峰表示的不是某个原子的位置，而是两个原子之间的向量峰。在一维的电子密度图上，假如在 x 处有一个峰，即表示了晶胞中的一个原子，而在一维的帕特逊图上，如果 u 处存在一个峰，则表示在相应的电子密度图中分别位于 x 处和 $x+u$ 处的两个峰(图6.1)。此处 u 既代表了原子之间的距离，又表示了方向，因此帕特逊图是向量图。这种图有如下几个特点：

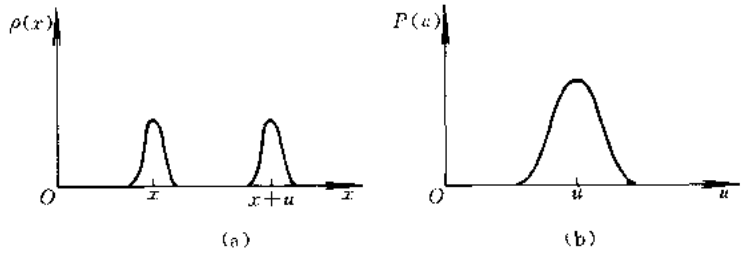


图 6.1

(a)示意在电子密度图上两个峰；(b)在相应的帕特逊图上出现的一个峰

(1) 若晶胞中有 N 个原子，则在向量图上有 N^2 个峰，其中 N 个是自身向量峰，即 $u=0$ 的峰，位于向量图的原点，而 N^2-N 个峰为非原点峰。

(2) 原点峰 $P_{(000)}$ 正比于 $\sum_{j=1}^{N} Z_j^2$，式中 Z_j 为第 j 个原子序数，$\sum_{j=1}^{N}$ 是对晶胞中全部 N 个原子的加和，而非原点峰正比于 $z_1 z_2$，z_1 和 z_2 为原子1和原子2的原子序数。由此可见，原子序数(z)很高的重原子之间的峰值，高于一般的轻原子之间和轻原子与重原子之间的峰值。

(3) 向量图是中心对称的，因原子1至原子2的向量($u_{1\to 2}$)和原子2至原子1的向量($u_{2\to 1}$)，方向相反而长度相等，例如一个三个原子的简单分子[图6.2(a)]，在相应的帕特逊向量图上将在原点周围出现中心对称的向量峰[图6.2(b)]。因此在向量图中可能出现的点群，仅具有中心对称性的11个点群。

根据上述晶胞中分子结构与它的帕特逊函数之间的关系，在原则上可以通过分析向量峰推算出原子的坐标，例如在 c 方向上具有一个二重轴的晶胞，若存在一个原子(x,y,z)，则必有二重轴联系起来的另一原子($\bar{x},\bar{y},z$)，见图6.3(a)。在相应的该两原子的 $P(uv0)$ 截面图上一定可以找到它们的向量峰(u,v)见图6.3(b)。在此 $u = x-\bar{x} = 2x$，$v = y-\bar{y} = 2y$，并在

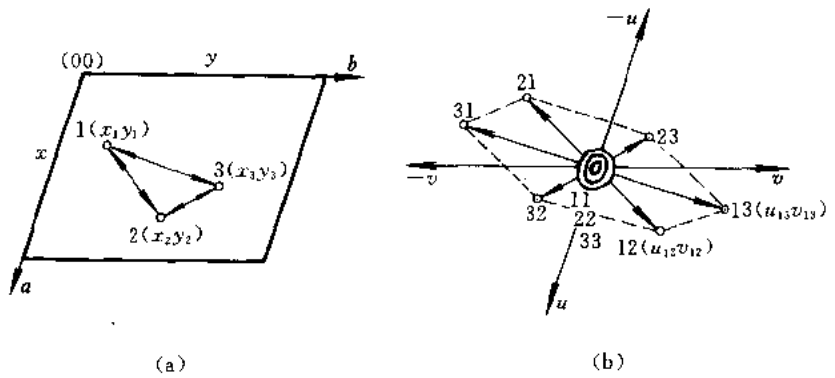


图6.2 分子结构与相应的向量峰

(a)在晶胞中一个三原子的分子结构;(b)在相应的向量图上出现的向量峰

截面图上可以读出 u,v 的具体数值,把读得的数值代入 u 与 x,v 与 y 的关系式,就可计算出 x,y 坐标值。当然这样测得的原子坐标是初步的,尚需进一步的精化。

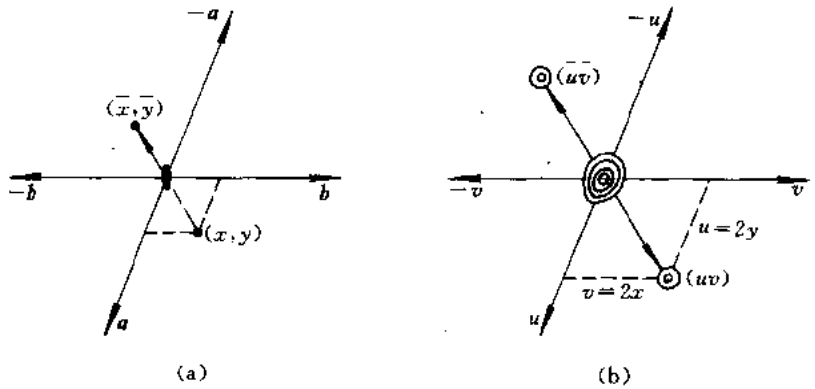


图6.3 晶胞中原子坐标与相应的向量峰坐标

(a)在晶胞中由2联系起来的两个原子;(b)在相应的 $P(uv0)$ 图中的向量峰

虽然对于生物大分子这样复杂的结构来说,由于在相应的向量图中峰的数目太多,重叠交盖很严重,根本无法通过分析直接辨认各原子间的向量峰,但是帕逊特函数在应用同晶置换法解决相角问题中,以及应用分子置换法确定两个相类似结构的工作中仍然是很重要的基本理论和技术。

6.2 同晶置换法

蛋白质等生物大分子主要由大量的 C,N,O,H 等轻原子组成,因而解决这类晶体结构分析的相角问题,主要是应用重原子同晶置换法。这种方法的基本思想是把散射本领大的重金属原子,如 Hg,Pb 等,引进到蛋白质晶体中作标识原子,然后设法解出这些数量较少的重原子在晶胞中的坐标,由这些坐标计算出重原子散射波在各衍射点的相角,再推引出蛋白质分子中每个原子散射的次生波在各级衍射中的位相。

蛋白质分子的晶体，除像小分子晶体一样，分子按空间格子排列以外，在分子之间还有充满结晶母液的小通道，通道宽度一般为2nm。在通道内溶剂分子与蛋白质分子表面的极性基团可相互作用而形成有规则的排列，因而外面的重原子通过小通道进入晶体内部以后，可以在一定位置上取代溶剂分子，形成重原子衍生物。而没有加重原子的晶体，相对于加重原子的衍生物而言称为母体。同晶置换是指两个或两个以上的晶体之间具有相同的空间群、相同的晶胞参数，以及绝大部分相同的原子分布，而仅在某些位置上被某种金属原子所取代。

两个晶体之间的同晶置换称为单对同晶置换，而一对以上的同晶置换称为多对同晶置换。蛋白质分子是非对映分子，因此总是结晶在非中心对称的空间群中，要解决晶体结构分析中的相角问题，一般讲都需要多对同晶。

6.2.1 同晶置换法测定相角的原理

前面已谈到晶体在某一衍射中的结构因子是晶胞中每个原子在该方向上散射 X 射线波的合成，同晶置换法测定相角的原理就是根据蛋白质重原子衍生物的某衍射的结构因子，可以由该衍射的母体结构因子和重原子结构因子的向量加和构成。其数学表示式为：

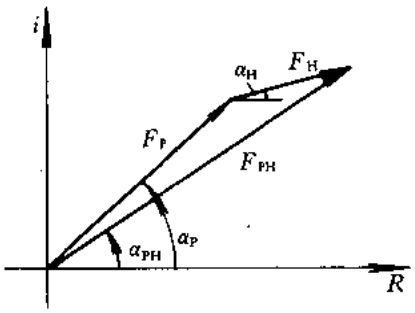


图6.4 在复数平面上 F_P，F_{PH} 和 F_H 的关系

$$F_{PH} = F_P + F_H$$

$$F_P = F_{PH} - F_H \tag{1}$$

F_P 是蛋白质母体的结构因子，F_{PH}是重原子衍生物的结构因子，F_H 是重原子的结构因子。(1)式亦可表达为

$$|F_P|e^{i\varphi_P} = |F_{PH}|e^{i\varphi_{PH}} - |F_H|e^{i\varphi_H}$$

φ_P，φ_{PH}，φ_H 分别表示蛋白质、衍生物和重原子次生波的相角。(1)式在复平面上的表示如图6.4。

如果令 R_1 和 R_2 分别代表母体和衍生物晶体之间相互置换的那部分原子，而令 L 代表两个晶体间相同的轻原子部分。则

$$F_P = F_L + F_{R_1}$$

$$F_{PH} = F_L + F_{R_2}$$

F_{R_1}和 F_{R_2}分别为相互置换的原子的结构因子，假定它们具有相同的坐标，则位相相同，仅结构振幅不同。由(1)式得

$$F_H = F_{PH} - F_P$$

则

$$\begin{aligned} F_H &= (F_L + F_{R_2}) - (F_L + F_{R_1}) \\ &= F_{R_2} - F_{R_1} \\ &= |F_{R_2}|e^{i\varphi_{R_2}} - |F_{R_1}|e^{i\varphi_{R_1}} \end{aligned}$$

由于 R_2，R_1的坐标相同，因此

$$F_H = (|F_{R_2}| - |F_{R_1}|)e^{i\varphi_H} = |F_H|e^{i\varphi_H}$$

而又知

$$F_H = \sum_{j=1}^{N} f_j e^{i2\pi(hx_j + ky_j + lz_j)}$$

所以假如精确知道了重原子的坐标 (x_j,y_j,z_j)，就可由上式计算出重原子在各级衍射中的位相，从而可知道 F_H 的大小和方向，然后根据公式(1)中衍生物、母体和重原子之间的向量关系就可设法推引出 F_P 的位相。因此要应用同晶置换法解决相角问题，就得首先测定重原子在晶胞中的位置。

6.2.2 重原子位置的测定

由公式(1)可得 $F_H=F_{PH}-F_P$，然而由于三者是向量关系，而 F_{PH}和 F_P 在此是未知的，因此不可能直接通过 F_{PH}和 F_P 来确定 F_H，但是由重原子衍生物和母体的衍射强度数据可计算出 F_{PH}和 F_P 的结构振幅值。假如能设法由这些实验的衍射数据引出 F_H 的振幅值($|F_H|$)，那就可应用帕特逊函数法来测定数量较少的重原子位置。

由图6.5的 F_{PH}、F_P 和 F_H 之间向量关系，可应用余弦定律得：

$$|F_H|^2=|F_{PH}|^2+|F_P|^2-2|F_{PH}||F_P|\cos\alpha$$

假定 F_{PH}与 F_P 之间的位相差较小，即 α 值较小，那末 $\cos\alpha\leqslant 1$，这时

$$|F_H|^2=|F_{PH}|^2+|F_P|^2-2|F_{PH}||F_P| \qquad (2)$$

而由母体和衍生物的衍射强度数据可计算出$|F_{PH}|$和$|F_P|$，并假定$|F_{PH}|-|F_P|$近似地等于$|F_H|$，则

$$(|F_{PH}|-|F_P|)^2=|F_{PH}|^2+|F_P|^2-2|F_{PH}||F_P| \qquad (3)$$

比较(2)(3)两式，可以说明用$(|F_{PH}|-|F_P|)^2$来代取$|F_H|^2$作帕特逊函数的系数还是合理的。这种以$(|F_{PH}|-|F_P|)^2$为系数的函数称为同晶差值帕特逊函数，具体形式如下：

$$P(uvw)=\frac{1}{V}\sum_{h}\sum_{+\infty}^{\infty}\sum_{k}\sum_{l}(|F_{PH}|-|F_P|)^2e^{-i2\pi(hu+kv+lw)}$$

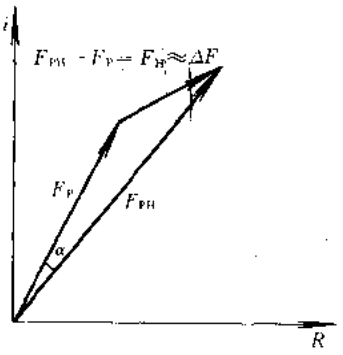


图6.5 示意 F_{PH}与 F_P 间的相角差

由于每个重原子的电子数目多，在晶胞中的数量又少，因此在帕特逊图上重原子之间的向量峰的峰值比一般峰高，而且数量较少，故易分析，从而有可能最终测定出重原子的坐标。假如已知重原子的坐标和占有率，就可从理论上计算每个衍射的 F_H，并将计算值与实验值进行比较，对重原子坐标采用数学方法作进一步修正，使理论计算值尽可能地接近实验值。

6.2.3 蛋白质母体的相角测定

6.2.3.1 单对同晶置换法

假如已经由一种重原子衍生物晶体的衍射强度数据计算出了结构因子的振幅$|F_{HP}|$的大小和由母体的衍射强度数据计算出了$|F_P|$的大小，并且已知 F_H 向量的大小和位相，就可通过作图法求出蛋白质母体的各衍射点的位相。

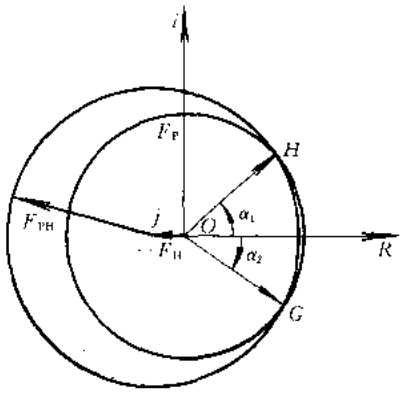


图6.6 作图法求相角示意图

首先按公式(1)在复平面原点 O，以 F_P 的大小为半径作一个圆，再在原点 O 作 $-F_H$，然后以向量 $-F_H$ 的终点 j 为圆心，以 F_{PH}的大小为半径画一个圆，一般两圆相交于两点(图6.6)。向

量$\overline{OH}$和$\overline{OG}$与极轴的夹角α_1和α_2即为F_P的两个可能的相角值，因而不能唯一地确定F_P的相角，一般需要另一对同晶置换的衍射数据，才能判断确定F_P的相角。但是如果F_P与F_H两向量共线，即以F_P和F_{PH}所作的两圆相切于一点，则有可能确定部分F_P的相角。

在什么情况下，才有可能出现上述F_P，F_H以及F_{PH}共线的结果呢，当晶体结构具有对称中心时，就会出现这种结果。当然蛋白质分子总是结晶在非中心对称的空间群里的，因为蛋白质分子都是非对映体，但是在某些非中心对称的空间群的某一特定方向的投影上可能出现对称中心，例如$P2_1$，C2空间群，沿$\underline{b}$方向上的投影，能出现对称中心。这时，$h0l$型的衍射点的相角，有可能利用单对同晶置换法来测定。

如果晶体结构具有对称中心，而且晶胞的原点处在对称中心上，那末晶胞的全部原子中，将有一半的坐标为x_j,y_j,z_j，另一半的原子坐标为$\bar{x}_j,\bar{y}_j,\bar{z}_j$。在此为讨论方便，假定晶胞中仅有两个碳原子，若一个在x,y,z，则另一个一定在$\bar{x},\bar{y},\bar{z}$。这时两个碳原子的结构因子为：

$$F_{hkl}=\sum_{j=1}^{2}f_c e^{i2\pi(hx_j+ky_j+lz_j)}$$

$$=\sum_{j=1}^{2}f_c[\cos 2\pi(hx_j+ky_j+lz_j)+i\sin 2\pi(hx_j+ky_j+lz_j)]$$

$$F_{hkl}=f_c[\cos 2\pi(hx+ky+lz)+\cos 2\pi(h\bar{x}+k\bar{y}+l\bar{z})$$

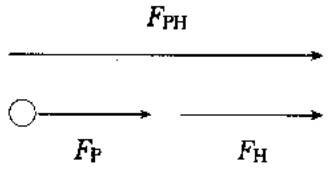
$$+i\sin 2\pi(hx+ky+lz)+i\sin 2\pi(h\bar{x}+k\bar{y}+l\bar{z})]$$

由于$\cos(-x)=\cos(x)$，$\sin(-x)=-\sin x$，则

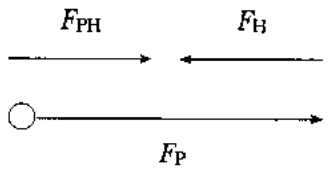
$$F_{hkl}=2f_c\cos 2\pi(hx+ky+lz)$$

这个结果说明由于结构为中心对称，结构因子的虚数项成对地相消了，仅存实数项。这样，结构因子的相角只可能为0或π，在复平面上仅分布在实轴上。

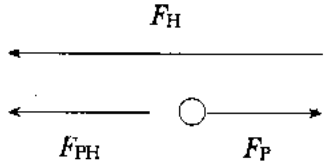
在二维中心对称区的F_P的相角可按下列几种情况确定，如果F_{PH}与F_P共线，按F_{PH}，F_F和F_H三者向量的关系，则当$F_{PH}>F_P$时，F_P与F_H的位相相同；



当$F_{PH}<F_P$时，F_P与F_H的位相相反；



当F_P非常小，而F_H很大时，F_P与F_H的位相相反，



当$F_{PH}\simeq F_P$时，就不可能确定F_P的位相。

6.2.3.2　多对同晶置换法

如果由第二对同晶衍生物的强度数据中引出了 F_{PH_2} 的大小和第二种重原子 F_{H_2} 的大小和位相，即已知 F_{H_2}，就可继续在图6.5上作图，解决那些单对同晶置换法无法确定的 F_P 的相角。

首先以原点 O 为起点作 $-F_{H_2}$，然后在 $-F_{H_2}$ 的末端，以 F_{PH_2} 的大小为半径作圆，该圆与 F_P 大小为半径的圆一般亦相截两点，假如为 H' 和 L。综合两次作图结果，由于第一个衍生物的 $\overrightarrow{OH}$ 向量和第二个衍生物的 $\overrightarrow{OH'}$ 向量的方向基本或完全一致，从而确定 α_1 为衍射 F_P 的相角，排除了 α_2 作为 F_P 的相角的可能性(图6.7)。假如 F_{H_2} 和 F_{H_1} 共线，则第二对同晶重原子衍生物就不能帮助挑出 α_1 和 α_2 中哪个是 F_P 的相角，那就需要第三对同晶重原子衍生物数据，才有可能准一确定该衍射点的相角。

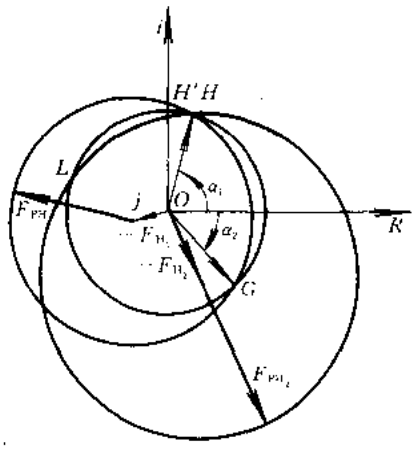


图6.7　在多对同晶置换法中应用作图法求相角的实例

6.3　反常散射法

原子散射因子通常是按原子中电子的空间分布情况来计算的，一般假定原子中的电子是不受束缚的，而事实上这种假定是不符合实际的，原子中的电子是受到束缚的，因此原子中的电子具有一个固有振动频率。当 X 射线入射到原子上时，它的散射能力与自由电子的散射能力相比较要有一个修正量。反常散射法就是利用这一性质来得到位相信息的。

在前面讨论原子散射因子时，是用实数量 f_j 表示，这是根据经典理论，假定入射线的频率远远大于原子的固有频率，而把原子中电子的束缚能量看成很小，近似于自由电子。在一般情况下，这种假设对大多数轻原子来说是正确的，但是对于重原子，特别是在入射线的波长靠近吸收限时，就不正确了。因为原子中的电子具有束缚能，其散射能力比自由电子大些，或者小些。因此散射因子就有一个修正量，这种效应称为反常散射。这时 f_j 成为复数形式，即为：

$$f_j = f_0 + \Delta f' + i\Delta f''$$

f_0 是通常的原子散射因子，$\Delta f'$ 是实部修正项，通常是负数，$\Delta f''$ 是虚部修正量，是正值。图6.8表示了具有反常散射原子的 f_0，$\Delta f'$ 和 $\Delta f''$ 之间的关系。

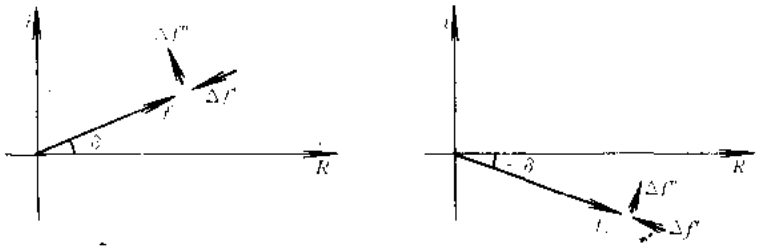


图6.8　在复平面上 f_0，$\Delta f'$ 和 $\Delta f''$ 间的关系

生物大分子晶体大多是由 C，H，O，N 等轻原子组成，对 X 射线，如 CuK$_\alpha$ 射线，吸收系数

是非常小的，可以忽略 $\Delta f''$，而仅仅考虑 $\Delta f'$。当蛋白质分子中存在重原子如 Fe 或加入重原子以后，由于重原子具有不可忽视的反常散射，它对结构因子 F_{hkl} 会产生影响。图6.9(a)表示了当不存在反常散射情况下，由于存在衍射花样强度分布中心对称定律，因此 $F_{hkl}=F_{\bar{h}\bar{k}\bar{l}}$，$\alpha_{hkl}=-\alpha_{\bar{h}\bar{k}\bar{l}}$，也就是说即使结构本身并不存在对称中心，但在衍射花样中强度将出现中心对称，$I_{hkl}=I_{\bar{h}\bar{k}\bar{l}}$。图6.9(b)表示了当蛋白质晶体结构中加入重原子以后，如果忽略反常散射时，结构因子 $F_{PH}(+)$ 仍然与 $F_{PH}(-)$ 相等，即弗里德耳定律仍然存在。图6.9(c)说明当重原子存在反常散射时，就破坏了弗里德耳定律，$F_{PH}(+)$ 就不等于 $F_{PH}(-)$。

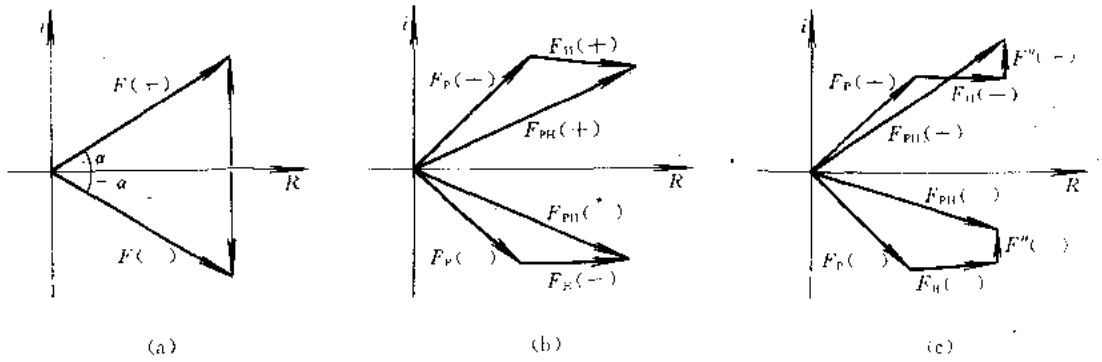


图 6.9

$F_P(+)$，$F_{PH}(+)$，$F_H(+)$分别代表 F_Phkl，$F_{PH}hkl$，F_Hhkl；$F_P(-)$，$F_{PH}(-)$，$F_H(-)$分别代表 $F_P\bar{h}\bar{k}\bar{l}$，$F_{PH}\bar{h}\bar{k}\bar{l}$，$F_H\bar{h}\bar{k}\bar{l}$。(a)在轻原子结构中不存在反常散射时结构因子间的关系；(b)在重原子衍生物中忽略反常散射对结构因子间的关系；(c)在重原子衍生物中存在反常散射时结构因子间的关系

在生物大分子晶体结构分析中应用反常散射法解决相角问题，主要是把反常散射法与单对同晶置换法相结合，解决 F_P 的相角的双解问题。例如在应用单对同晶置换法解决相角时，一般出现双解，不能唯一测定 F_P 的相角[图6.10(a)]，而若没有同晶母体仅单独应用反常散射效应法解决 F_P 的相角，亦往往出现双解[图6.10(b)]，图中以 $F_{PH}(+)$，$F_{PH}(-)$为半径的两圆相截在 M 和 H，F_P 的可能相角为$\overrightarrow{OH}$和$\overrightarrow{OM}$两向量的相角，不能唯一测定。但是，假如已收集了该同晶重原子衍生物的反常散射数据，即已有 $F_{PH}(+)$，$F_{PH}(-)$和 $F''_H(+)$，$F''_H(-)$的数值，就可把两种方法结合起来，有可能唯一地确定 F_P 的相角。比较图6.10(a)和(b)，在(b)中 F_P，$F_{PH}(+)$，$F_{PH}(-)$三个圆都相截在 H，由此可得出结论，向量$\overrightarrow{OH}$的相角 α_1即为某一衍射 F_P 的相角。

前面讨论了同晶置换法和应用反常散射效应的一般原理，而在实际工作中要复杂得多，大体有如下几个步骤。

(1) 重原子衍生物的制备。如前述制备重原子衍生物就是要把重原子引进到晶体中去而不破坏晶格和生物大分子的结构。一般采取尝试的方法，不同的生物大分子晶体使用不同的方法，无一固定的程序。首先要详细查阅有关的文献，了解有关的晶体结构的细节，提出一个较合理的制备途径。例如根据一级结构和高级结构的有关资料，分析哪些氨基酸可能埋在蛋白质分子的核心，哪些可能处在晶格中分子接触的位置，这些部位重原子将是不可进入的，而那些处在分子表面的氨基酸将有可能提供重原子的结合部位。如果蛋白质含有金属原子如锌或钙，可

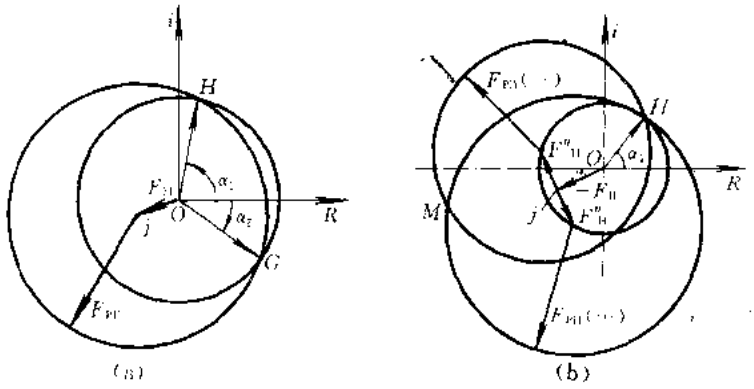


图6.10 单对同晶置换与反常散射相结合测定 F_P 的相角

(a)单对同晶置换法中 F_P 的相角出现 α_1 和 α_2 两个可能；(b)应用反常散射帮助确定 F_P 的相角为 α_1

以考虑用更重的金属原子去取代，并根据金属离子半径的大小，推测可以应用哪些金属原子去取代，而不会破坏蛋白质分子的结构。金属与大分子结合的方式亦要考虑，与羧酸基、咪唑基结合的锌、钙离子较易去掉而不破坏分子的高级结构。

当前制备重原子同晶衍生物最普遍、最有效的方法是浸泡法。这种方法就是把蛋白质晶体浸泡在含有重原子的结晶母液中，重原子沿着晶体内充满溶剂的小通道扩散到蛋白质晶体中，与蛋白质侧链的功能基团相互作用，因而可以得到合适的重原子衍生物。虽然这种作用往往缺乏特殊性，但是晶体学家已经积累了许多经验。另外，如果研究的对象是一个酶，可以根据活动中心的性质，合成酶底物的重原子类似物，使其与酶结合而制备重原子衍生物。

(2) 测量蛋白质母体和衍生物晶体的衍射强度。

(3) 强度数据的还原和修正。

(4) 重原子位置的测定。

(5) 相角的测定。

迄今为止，同晶置换法仍然是蛋白质晶体结构分析的主要方法，在今后仍将起着重要的作用。我国在1972年完成的0.25nm分辨率的胰岛素晶体结构分析，主要也是应用多对同晶置换法。当时获得了三对(汞胰岛素、铅胰岛素和汞铅胰岛素)质量很高的同晶衍生物，解决了相角问题，得到了质量很好的电子密度图，使后来的整个结构分析工作达到了世界先进水平。

6.4 分子置换法

在第三章讨论衍射花样时，已谈到分子在晶体中的排列方式决定衍射花样中衍射点的分布方式，而分子结构本身的特点决定衍射强度分布的特点，即是说在分子排列以及分子结构上完全相同或类似的晶体在衍射花样上亦必然出现相同或类似的特征，这种衍射花样上的共同或类似的特征同样会反映在它们的帕特逊向量图上。因此，假如有两个在分子结构上相同或相似但具有不同晶型的晶体，而其中一个晶体结构已测定，那么根据上述道理从原则上可以设法由已知晶体结构来推引出未知晶体结构中相同或类似分子在不同晶胞中的取向和位置，从而得到分子结构的初始模型。解决这一类结构问题的方法称为分子置换法。这种方法首先计算已知的模型分子和待测分子的自身帕特逊函数，然后把两者放到模型晶胞中通过旋转操作，确定

待测分子的取向。第二步是通过比较模型晶体与待测晶体分子间的交叉帕特逊向量，得到有关平移向量的信息，然后进行平移操作确定待测分子在晶胞中的正确位置。最后根据部分或全部结构模型提取位相信息；如果待测分子不但与已知晶体结构的分子在三维结构上类似，而且两者是同晶型，则可应用差值傅里叶合成方式测定晶体结构。

6.4.1 分子置换法在晶体结构测定中的主要应用

(1) 测定非晶体学对称性：很多球蛋白分子是寡聚蛋白，即由一个以上的亚单位彼此以非共价键相联系聚合而成的蛋白质分子。亚单位具有自己的三级结构，是蛋白质分子结构的最小共价单位。在分子中各亚单位一般按一定的对称轴在空间排布，当这类蛋白质分子形成晶体时，具有对称轴的分子可以坐落在晶体的对称轴上，这样每一个晶体的不对称单位仅有一个亚单位，但是相当多的情况是分子不坐落在晶体的对称轴上，因此每个不对称单位是由1个亚基或2个以上的亚基寡聚体分子构成。另一种情况是具有一个亚单位的蛋白质分子亦可彼此聚合，形成一个不对称单位，例如三方胰岛素晶体由二聚体分子构成一个不对称单位，两个完全相同的分子由近似的二重轴联系。这种在不对称单位内联系各亚单位或分子之间的对称轴称为非晶体学对称轴，或称为局部对称轴。这种由非晶体学对称轴联系起来的亚单位或分子之间，仅取向和位置不同，而每个亚单位或分子的三级结构是相同的，或者是非常近似的。

(2) 利用已知分子结构测定多晶型蛋白质分子的结构：在蛋白质晶体学的研究中，经常出现多晶型的现象，有时当晶体的生长条件略有改变时，例如 pH 的变化；溶剂或沉淀剂的改变，同一种蛋白质分子可以形成不同的晶型，即结晶在不同的空间群中。这种不同晶型的结构之间仅仅分子的堆积方式不同，而分子的结构基本相同。例如胰岛素分子可结晶在 R_3 空间群中，亦可结晶在 $P2_12_12_1$ 空间群中。

(3) 利用已知分子结构测定同源蛋白分子的结构：在蛋白质中有一类是同源蛋白，它们由共同的较原始的分子进化而来，这类蛋白质分子在一级结构上有共同的保守区，特别在三级结构上有绝大部分是相同的结构。例如血红蛋白和肌红蛋白；糜蛋白酶、胰蛋白酶和弹性蛋白酶都是同源的。但是在晶体生长时同源蛋白往往结晶在完全不同的空间群中，这种不同的空间群的晶体结构之间仅仅分子的堆积方式不同，而分子的三级结构基本相似。

(4) 研究酶和抑制剂的复合物中酶与抑制剂之间的相互作用：一个酶分子在不同条件下可以与不同的抑制剂形成复合物。在这些不同复合物中，酶分子的结构是相同的，仅抑制剂的分子结构不同，因此可以通过已知酶分子的结构，应用分子置换法来测定抑制剂的结构，并研究酶与抑制剂的相互作用。

6.4.2 分子置换法的基本原理

假如已知结构的分子 M(具有一个或一个以上的亚单位)与待测晶体中的未知结构分子 M′有基本相似的三级结构，可以把分子 M 作为 M′的模型分子，采用该模型分子按一定的原则构建模型晶胞。设分子 M 中各个原子在模型晶胞中的位置向量为 r_j，$j=1,2,\cdots,N$，N 为 M 分子中的原子数，未知分子 M′中相应原子在待测晶体的晶胞中的位置向量为 γ_j，则：

$$\gamma_j = R\gamma_j + d$$

式中 R 是一个旋转矩阵，d 为平移向量。通过 R 矩阵的旋转操作，可以使模型分子 M 在晶胞中的取向与 M′分子一致，然后通过平移向量 d 把分子 M 移到晶胞中的正确位置，从而得到未知晶体的初步模型。在对取向角和平移向量作进一步地修正之后，就可采取加权的傅里叶合成法和差值傅里叶合成法计算出待测分子的电子密度图，然后对电子密度图进行解释。应用同晶置

换法解决相角问题，就不需要制备重原子衍生物，这样可节省人力和时间。

应用分子置换法时，需要具有与未知分子结构相似的已知结构，以及正确求出 R 和 d，现在较为广泛应用的是通过比较模型晶体和未知结构的晶体的帕特逊函数来求解 R 和 d。

6.4.2.1 旋转函数

假如把已知结构的模型分子按任意取向和位置安放在模型晶胞中[图6.11(a)]。假定待测晶体的晶胞如图6.12(a)所示。为了方便起见，假设两种晶体均属 P_1 空间群，并分别计算两个

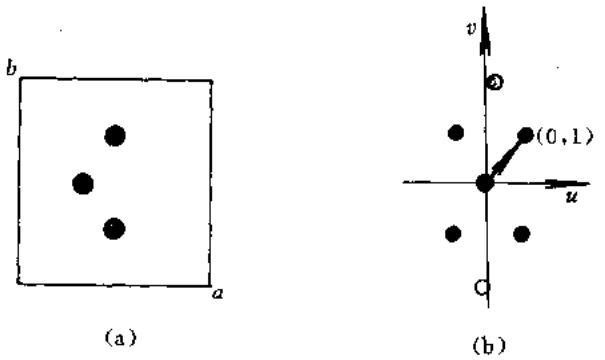


图6.11 模型分子和模型分子的自身向量

(a)模型晶胞中的分子；(b)模型分子的自身向量图

晶体的帕特逊函数，图6.11(b)和图6.12(b)分别给出了模型晶体和待测晶体的自身向量图，即分子内部各原子之间的向量图。自身向量的取向直接与分子在晶胞中的取向有关。仔细观察

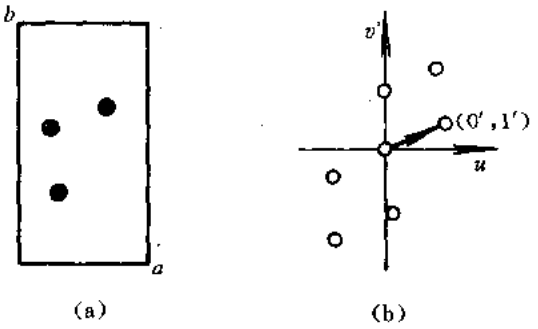


图6.12 待测分子和待测分子的自身向量

(a)待测晶胞中的分子；(b)待测分子的自身向量图

和比较两个晶体的自身向量图就会发现，这两套图中的相应向量峰，例如(0,1)与(0′,1′)间相差的角度，恰好等于模型分子与待测分子取向相差的角度。如果将模型分子沿顺时针方向旋转25°左右，它的取向便和待测分子在晶胞中的取向相同，此时，模型分子和待测分子的向量峰的取向也相同。通过上面的讨论可以看出，通过模型分子和待测分子的自身向量图的比较，可以求出两个分子取向间的相对角度差，通过某种旋转操作，可以使模型分子在模型晶胞中的取向与待测分子在晶胞中的取向一致。

罗斯曼(G. M. Rossmann)和布洛(M. D. Blow)等人根据上述基本原理提出了一种旋转函

数 R，来确定两分子取向间的角度差。

$$R(\theta_1,\theta_2,\theta_3)=\int_u P_2(x')P_1(x)\mathrm{d}x$$

式中 $P_1(x)$和 $P_2(x')$分别代表模型晶体和待测晶体的帕特逊函数，U 为积分体积；$\theta_1,\theta_2,\theta_3$是在三维空间中旋转操作的三个欧拉角。假如 $P_1(x)$和 $P_2(x')$分别为一套模型分子和待测分子的自身帕特逊向量群，则称 R 为交叉旋转函数。当不断改变旋转角$(\theta_1,\theta_2,\theta_3)$致使两个分子的取向逐步达到一致时，$R$ 函数应当出现极大值。这时的欧拉角，就是待测分子相对于已知分子取向的角度差。根据已求得的角度差$(\theta_1,\theta_2,\theta_3)$对已知分子进行旋转操作，使它的取向与待测分子的取向完全一致，这时已知分子的取向就是待测分子在晶胞中的绝对取向。如果 $P_1(x)$和 $P_2(x')$均为待测分子的自身帕特逊向量群，则称 R 函数为自身旋转函数，当不断改变旋转角$(\theta_1,\theta_2,\theta_3)$致使两个待测分子中的等同部分，例如亚基取向完全一致时，R 函数亦应出现极大值，由相应的旋转角就可推测出待测分子的非晶体学对称性。

6.4.2.2　平移函数

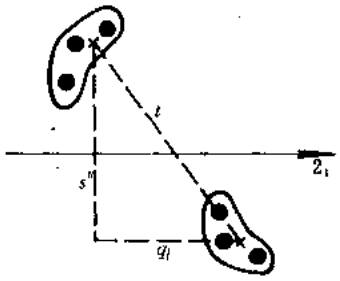
在确定了模型分子即待测分子的绝对取向之后，下一步工作就是要设法确定平移向量 d，然后通过平移把已经有了正确取向的模型分子放在晶胞中的正确位置，即与待测分子完全一致的位置。

为便于讨论如何确定平移向量 d，假设在模型晶体中存在一个包含平移动作的对称元素 2_1轴，由该2_1轴联系起来的两个模型分子的质心之间的向量 t 为已知[图6.13(a)]。而亦假定在待测晶体中由2_1轴联系起来的、两个待测分子的质心之间的向量为 t'，而 t' 是未知的，因此是待测的参数[图6.14(a)]。图中×表示分子的质心，q 向量和2_1轴平行。模型晶体和待测晶体的帕特逊向量图分别在图6.13(b)和图6.14(b)给出。图中△表示分子1内各原子间的自身向量峰，○表示分子2的自身向量峰，●表示分子1和分子2的原子间的交叉向量峰。由图可以看出，两个晶体的交叉向量峰的取向是一致的，但位置与向量 t 和 t'有关。改变模型晶体中连结分子1和分子2之间的向量 t 至使它刚好等于待测晶体中的向量 t'时，两张帕特逊向量图应当一致。

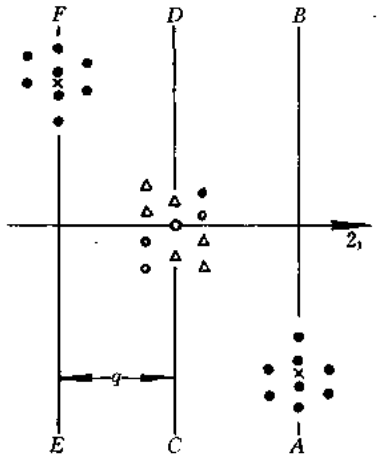
根据上述原理，克罗萨尔(A. R. Crowth)等人提出了平移函数 $T(t)$：

$$T(t)=\int P(u)P_{\mathrm{M}}(v,t)\mathrm{d}u$$

式中 $P(u)$为待测晶体的帕特逊函数，它是由待测晶体的衍射强度数据引出的结构振幅实验值计算得到的，$P_{\mathrm{M}}(u,t)$为模型晶体由指定的对称元素联系起来的、两个模型分子之间的交叉帕特逊函数。它是由已知模型晶体的结构参数从理论上计算的结构振幅值计算得到的交叉帕特逊函数。当连结两个模型分子质心之间的向量 t 刚好等于待测晶体中的两个分子之间的向量 t'时，$T(t)$函数将有一个极大值。用这种方法最终可求出平移向量 d，然后将模型分子移至正确的位置，即与待测分子在晶胞中的取向和位置基本相同。对取向角和平移量作进一步的修正之后，就可用已知结构的模型分子来计算各衍射点位相，并与待测分子的衍射强度实验值相结合，采用傅里叶合成法计算电子密度图。[10][14]

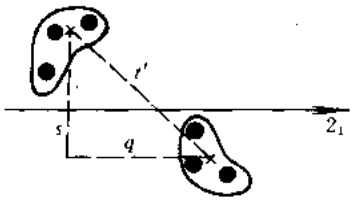


(a) 两个模型分子的相互关系

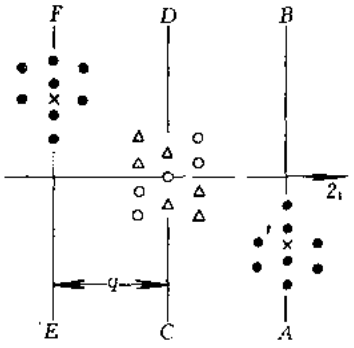


(b) 向量图

图6.13 模型晶体中由2_1轴联系起来的两个模型分子及其向量图



(a) 两个待测分子的相互关系



(b) 向量图

图6.14 待测晶体中由2_1轴联系起来的两个分子及其向量图

附录 65个非对映空间群的衍射群和系统消光判据

说明：

(1) 表中衍射符号应包括劳埃群和消光符号，现均略去劳埃点群部分，例如单斜晶系中空间群 $P2_1$的衍射符号应为$2/m \cdot P \cdot 2_1$。

(2) 在衍射指标中 $i=-(h+k)$；对 R 点群 $h0\bar{h}l$ 的消光判据也适用于 $h\bar{h}0l$ 及$0h\,\bar{h}l$ 类型。

表1 三斜晶系 劳埃点群 $C_i-\bar{1}$

序号	衍射符号	系统消光判据			空间群	
		hkl			1	
1	$P\bar{1}$				$P1$	

表2　单斜晶系　劳埃点群　$C_{2h}-2/m$

序号	衍射符号	系统消光判据			空间群	
		hkl	h0l	0k0	2	
1	P…				P2	
2	P·2_1·			2n+1	$P2_1$	
3	C…	h+k=2n+1	h+k=2n+1	2n+1	C2	

表3　正交晶系　劳厄点群　$D_{2h}-mmm$

序号	衍射符号	系统消光判据							空间群
		hkl	0*kl*	*h*0*l*	*hk*0	*h*00	0*k*0	00*l*	222
1	*P*…								*P*222
2	*P*··2_1							2*n*+1	*P*222$_1$
3	*P*$2_1$$2_1$·					2*n*+1	2*n*+1		*P*$2_1$$2_1$2
4	*P*$2_1$$2_1$$2_1$					2*n*+1	2*n*+1	2*n*+1	*P*$2_1$$2_1$$2_1$
5	*C*…	*h*+*k*=2*n*+1	*h*+*k*=2*n*+1	*h*+*k*=2*n*+1	*h*+*k*+*l*=2*n*+1	2*n*+1	2*n*+1		*C*222
6	*C*··2_1	*h*+*k*=2*n*+1	*h*+*k*=2*n*+1	*h*+*k*=2*n*+1	*h*+*k*+*l*=2*n*+1	2*n*+1	2*n*+1	2*n*+1	*C*222$_1$
7	*I*…	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	2*n*+1	2*n*+1	2*n*+1	*I*222 *I*$2_1$$2_1$$2_1$
8	*F*…	*h*,*k*,*l* 不为全奇或全偶				2*n*+1	2*n*+1	2*n*+1	*F*222

表4　四方晶系　劳厄点群　$C_{4h}-4/m$

序号	衍射符号	系统消光判据							空间群
		hkl	*hk*0	0*kl*	*hhl*	00*l*	0*k*0	*h*00	4
1	*P*…								*P*4
2	*P*4_2··					2*n*+1			*P*4_2
3	*P*4_1··					不为4的倍数			*P*4_1 *P*4_3
4	*I*…	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	2*n*+1	2*n*+1		*I*$_4$
5	*I*4_1··	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	*h*+*k*+*l*=2*n*+1	不为4的倍数	2*n*+1		*I*4_1

表5　四方晶系　劳埃点群　$D_{4h}-4/mmm$

序号	衍射符号	系统消光判据							空间群
		hkl	$hk0$	$0kl$	hhl	$00l$	$0k0$	$h00$	422
1	$P\cdots$								$P422$
2	$P4_2\cdot\cdot$					$2n+1$			$P4_222$
3	$P4_1\cdot\cdot$					不为4的倍数			$P4_122$ $P4_322$
4	$P\cdot2_1\cdot$						$2n+1$		$P42_12$
5	$P4_22_1\cdot$					$2n+1$	$2n+1$		$P4_22_12$
6	$P4_12_1\cdot$					不为4的倍数	$2n+1$		$P4_12_12$ $P4_32_12$
7	$I\cdots$	$h+k+l=2n+1$	$h+k+l=2n+1$	$h+k+l=2n+1$	$h+k+l=2n+1$	$2n+1$	$2n+1$		$I422$
8	$I4_1\cdot\cdot$	$h+k+l=2n+1$	$h+k+l=2n+1$	$h+k+l=2n+1$	$h+k+l=2n+1$	不为4的倍数	$2n+1$		$I4_122$

表6　三方晶系　劳埃点群　$C_{3i}-\bar{3}$

序号	衍射符号	系统消光判据				空间群
		$hkil$	$h0\bar{h}l$	$hh\,\overline{2h}l$	$000l$	3
1	$P\cdots$					$P3$
2	$P3_1\cdot\cdot$				不为3的倍数	$P3_1$ $P3_2$
3	R	$-h+k+l$ 不为3的倍数			不为3的倍数	$R3$

表7　三方晶系　劳埃点群　$D_{3d}-\bar{3}m$

序号	衍射符号	系统消光判据				空间群
		$hkil$	$h0\bar{h}l$	$hh\,\overline{2h}l$	$000l$	32
1a	$P\cdot1\cdot$					$P312$
1b	$P\cdot\cdot1$					$P321$
2a	$P3_11\cdot$				不为3的倍数	$P3_112$　$P3_212$
2b	$P3_1\cdot1$				不为3的倍数	$P3_121$　$P3_221$
3	$R\cdots$	$-h+k+l$ 不为3的倍数者消光			不为3的倍数	$R32$

表8　六方晶系　劳埃点群　$C_{6h}-6/m$

序号	衍射符号	系统消光判据				空间群
		$hkil$	$h0\bar{h}l$	$hh\,\overline{2h}l$	$000l$	6
4	$P\cdots$					$P6$
5	$P6_3\cdot\cdot$				$2n+1$	$P6_3$
6	$P6_2\cdot\cdot$				不为3的倍数	$P6_2$ $P6_4$
7	$P6_1\cdot\cdot$			不为6的倍数		$P6_1$ $P6_5$

表9　六方晶系　劳埃点群　$D_{6h}-6/mmm$

序号	衍射符号	系统消光判据				空间群
		$hkil$	$h0\bar{h}l$	$hh\,\overline{2h}l$	$000l$	622
1	$P\cdots$					$P622$
2	$P6_3\cdot\cdot$				$2n+1$	$P6_322$
3	$P6_2\cdot\cdot$				不为3的倍数	$P6_222$ $P6_422$
4	$P6_1\cdot\cdot$			不为6的倍数		$P6_122$ $P6_522$

表10　立方晶系　劳埃点群　$T_h-m\bar{3}$

序号	衍射符号	系统消光判据				空间群
		hkl	$0kl$	hhl	$00l$	23
1	$P\cdots$					$P23$
2	$P2_1\cdot\cdot$				$2n+1$	$P2_13$
3	$I\cdots$	$h+k+l$ 为奇数者消光			$2n+1$	$I23$ $I2_13$
4	$F\cdot\cdot$	h,k,l 不为全奇或全偶者消光			$2n+1$	$F23$

表11　立方晶系　劳埃点群　$O_h-m\bar{3}m$

序号	衍射符号	系统消光判据				空间群
		hkl	$0kl$	hhl	$00l$	432
1	$P\cdots$					$P432$
2	$P4_2\cdot\cdot$				$2n+1$	$P4_232$
3	$P4_1\cdot\cdot$				不为4的倍数	$P4_132$ $P4_332$
4	$I\cdots$	$h+k+l=$奇数者消光			$2n+1$	$I432$
5	$I4_1\cdot\cdot$	$h+k+l=$奇数者消光			不为4的倍数	$I4_132$
6	$F\cdots$	h,k,l 不为全奇或全偶者消光			$2n+1$	$F432$
7	$F4_1\cdot\cdot$	h,k,l 不为全奇或全偶者消光			不为4的倍数	$F4_132$

第二篇　生物大分子的立体结构

自从X射线衍射分析方法建立以来，生物来源物质的三维空间结构就一直是人们注意的焦点。在最早的20年里，拍摄到的天然纤维——蚕丝和纤维素的X射线衍射照片证明了它们的结构具有高度的有序性。但由于这些物质的结构十分复杂，而且当时缺少综合的化学数据，所以在很长时期内不能详细地被解释。然而正是这些研究对象特殊的复杂性刺激了X射线衍射实验方法和晶体学结构分析理论以及计算技术的改进和发展。由X射线衍射分析所得到的成果和所提供的有价值的资料已经使生物学的发展发生了根本的变化。现在，已经获得了大量不同的生物大分子包括蛋白质、核酸、病毒、多糖等等的原子坐标数据，并逐渐认识了它们的三维结构的特点和规律。一个新的学科——结构生物学已经形成。可以说X射线衍射技术是结构生物学研究的最有效的工具之一，它仍然是目前能在原子水平上提供生物大分子三维结构资料的最有效的技术。

到现在为止，几乎所有重要的生物分子都已用X射线衍射方法在直到原子分辨率水平上进行了研究：

氨基酸、核苷酸、单糖的分子能以有序的晶体型式存在，并以氢键为特征，晶胞的大小一般为0.5—2nm。这些简单的分子中的原子数目可达10^2，它们的极大的重要性在于它们是几乎所有生物结构的建造单元。目前，已经测定了所有这些分子以及大量的衍生物的晶体结构。

肽、类固醇、激素、维生素以及脂类等的一个分子中所含的原子数可达10^2—10^3，从生物学的观点看仍是很小的分子。但对于X射线晶体学来说，已经是非常复杂的了。这些分子在体内的生物学过程的控制方面通常起着重要的作用。它们能以晶体或液晶等有序性型式存在，已经得到了一些关于这些分子和它们的晶体结构的资料。它们的晶胞大小一般为1—3nm。

纤维状蛋白质、球状蛋白质、核酸和多糖等生物大分子的一个分子中或一个亚基中含有的原子数目可达10^2—10^5。这些生物大分子是现代晶体学研究最活跃的领域。它们能以纤维结构、层状结构或晶体型式存在，晶胞的大小或链的周期一般可从几个纳米直到20纳米。

生物膜、核蛋白、染色体、核糖体及病毒等是更为复杂的生物学对象。它们是由大量生物分子如蛋白质、核酸、脂类或糖类等结合或组装而成的。在一个分子或一个亚基中的原子的数目可达10^2—10^8。它们能以纤维状结构、层状结构或晶体等有序性的型式存在，晶胞的大小或链的周期可从几个纳米直到几十甚至几百纳米。它们是最近十多年来晶体学研究的最新和最有吸引力的领域。

第七章　蛋白质的立体结构

蛋白质是生物体中最重要的生物大分子之一，它对所有生物学过程起着决定性的作用。

蛋白质分子是由各种氨基酸以肽键相连接而成的大分子，它们的分子结构可以被描述为一级、二级、三级和四级结构。一级结构是指肽链中的氨基酸顺序，即化学分子式。肽键是一种共价键，因此蛋白质的一级结构是一种共价结构。此外，蛋白质的重要特点是它们具有十分确定的立体结构（或高级结构），包括二级结构——多肽链排列成一个由氢键系统所稳定的确定的各种构象；三级结构——具有二级结构的多肽链在空间盘绕折叠成固定的复杂的构象；以及四级结构——蛋白质的若干个具有三级结构的亚基缔合并组装成一定的空间关系。蛋白质的高级结构与功能的关系已成为蛋白质研究的中心问题之一。

7.1　氨基酸和多肽链的立体结构和构象角的定义[20]

蛋白质是由20种具有不同侧链基团 R 的 L 构型的 α-氨基酸所组成的。它们的立体结构示于图7.1和图7.2。除甘氨酸以外，其他19种氨基酸都有一个不对称的 C_α 原子，此外，异亮氨酸和苏氨酸的侧链中还各有另一个不对称 C 原子。在蛋白质中，这20种氨基酸所具有的共同的 L 构型限制了肽链的构象和折叠的可能形式，而20种不同的侧链及其在蛋白质序列中的位置对每个蛋白质独特的三维空间结构又具有各自特殊的决定作用。

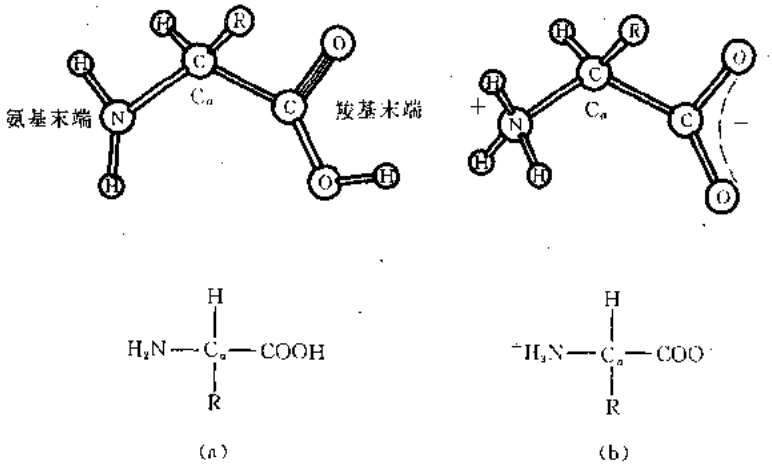


图7.1　α-L-氨基酸的立体结构

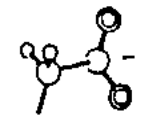
(a)不带电荷的型式；(b)两性离子的型式。R 为氨基酸的不同侧链

氨基酸以相邻氨基酸的氨基和羧基脱水生成的肽键（—CO—NH—）相连而形成多肽链。图7.3标出了肽键的标准键长和键角，图7.4表示一个完全伸展的多肽链的立体结构。肽键是刚性的平面，称为酰胺平面，其中 N 原子与羧基 C 原子之间的键具有部分双键性质，不能旋转；

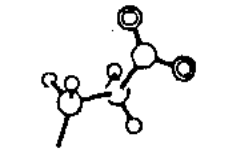
而刚性肽键两侧的 C_α—C 和 C_α—N 单键则可以自由旋转。由于一条多肽链上的所有单键都可旋转，因而肽链可能形成无数种不同的构象。因为蛋白质是由具有手性的 L-氨基酸缩合而成的，所以多肽链的最稳定构象不是完全伸展的构象，而是在三维空间中折叠着的某种构象。

位于外部的：

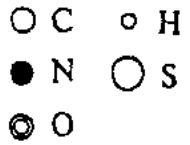
(a) 酸性的



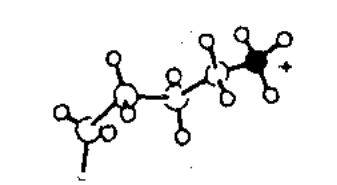
天冬氨酸(Asp 或 D)



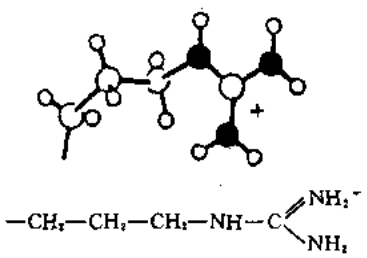
谷氨酸(Glu 或 E)



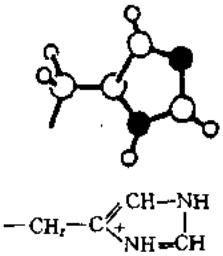
(b) 碱性的



赖氨酸(Lys 或 K)

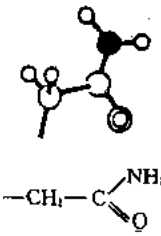


精氨酸(Arg 或 R)

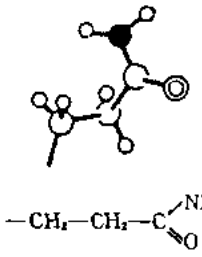


组氨酸(His 或 H)

(c) 中性的

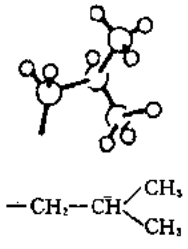


天冬酰胺(Asn 或 N)

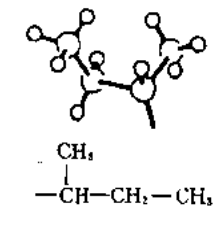


谷酰胺(Gln 或 Q)

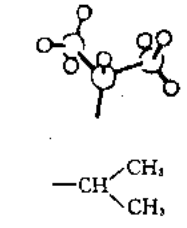
位于内部的：



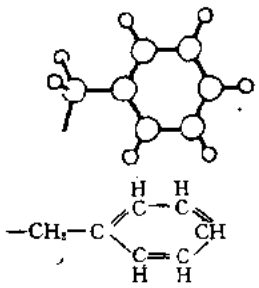
亮氨酸(Leu 或 L)



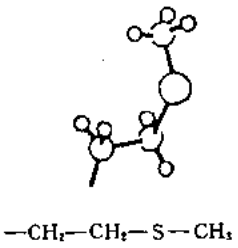
异亮氨酸(Ile 或 I)



缬氨酸(Val 或 V)



苯丙氨酸(Phe 或 F)

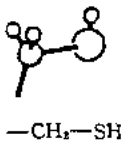


蛋氨酸(Met 或 M)

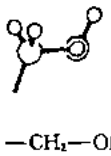
位于内外部两可的：

甘氨酸(Gly 或 G)

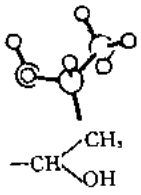
丙氨酸(Ala 或 A)



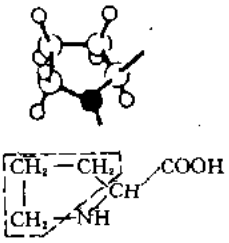
半胱氨酸(Cys 或 C)



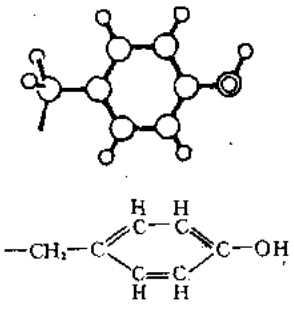
丝氨酸(Ser 或 S)



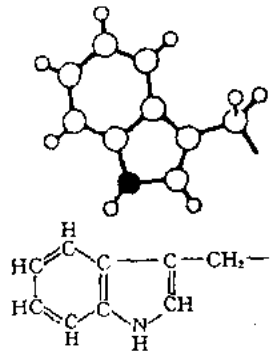
苏氨酸(Thr 或 T)



脯氨酸(Pro 或 P)
(虚线框内的为侧链)



酪氨酸(Tyr 或 Y)



色氨酸(Trp 或 W)

图7.2　20种氨基酸的侧链结构

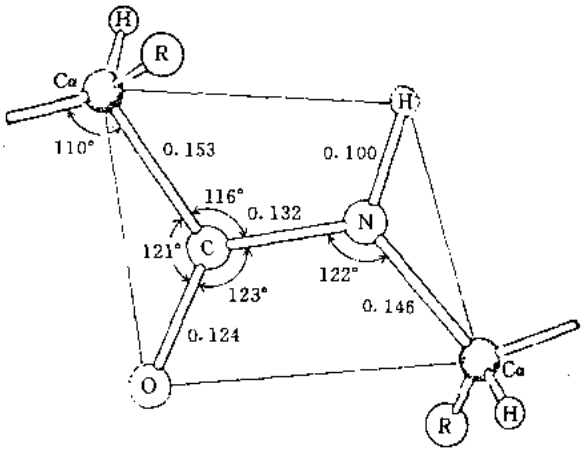


图7.3 肽键的标准键长(以 nm 为单位)和键角

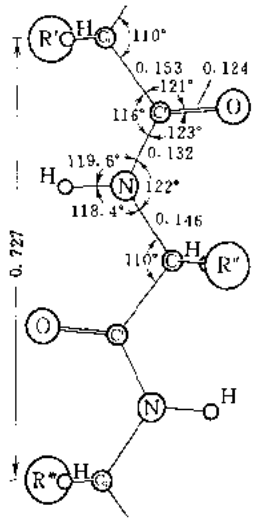


图7.4 一个完全伸展的多肽链结构示意图
其中键长以 nm 为单位

为了精确地描述蛋白质的三维构象，用扭角(或双面角)来定义各个构象角。如图7.5所示，所谓扭角，涉及4个原子 A,B,C 和 D,定义为平面 ABC 与平面 BCD 之间的夹角，以 θ 表示。沿着中央两个原子连线的方向$\overrightarrow{BC}$(或$\overrightarrow{CB}$,两种方向均可)观看时，以近端的一个键 A—B(或 C—D)作为标准参考，定义为0°，若远端的键 C—D(或 A—B)相对于0°为顺时针方向旋转时，定义 θ 为正值；若为反时针方向，则定义 θ 为负值；如果远端键与近端键投影重合形成顺式构象，定义为 $\theta=0°$；若为反式构象，则 $\theta=180°$，并规定 $-180°<\theta\leqslant 180°$。

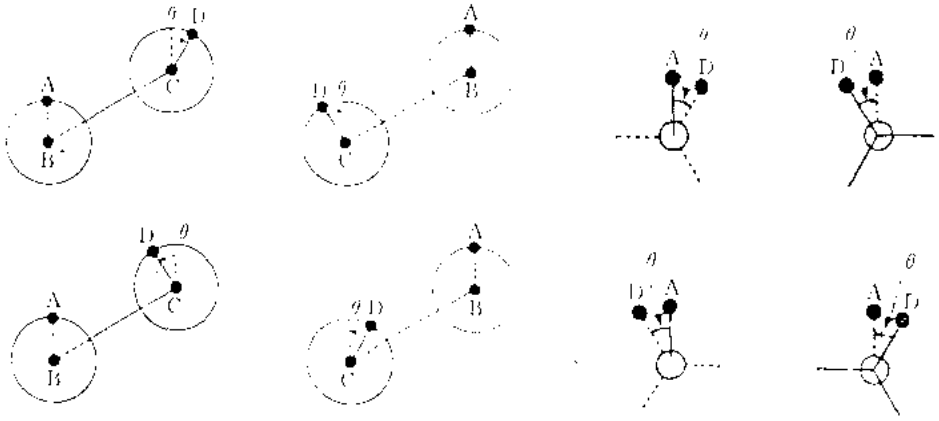


图7.5 扭角的定义

对于多肽链主链的各个扭角定义为：沿着主链的氨基酸序列的方向，将围绕着 N—C_α 键、C_α—C 键以及 C—N 键的旋转各自所涉及到的4个原子的扭角，分别以希腊字母 ϕ,ψ 和 ω 表

示；用同样的规则，将侧链的扭角从靠近主链的一端开始依次分别用 x_1, x_2, x_3…等等表示，如图7.6所示。其中，由于肽键是一个刚性平面而且通常是反式构象，因此几乎所有多肽链的 ω 都接近于±180°，不过有时甘氨酸或脯氨酸可形成顺式肽键，即 ω 接近于0°。所以，另外两个扭角 ϕ 和 ψ 的旋转实际上是所有蛋白质构象变化的来源，是蛋白质立体结构的重要参数。但是，

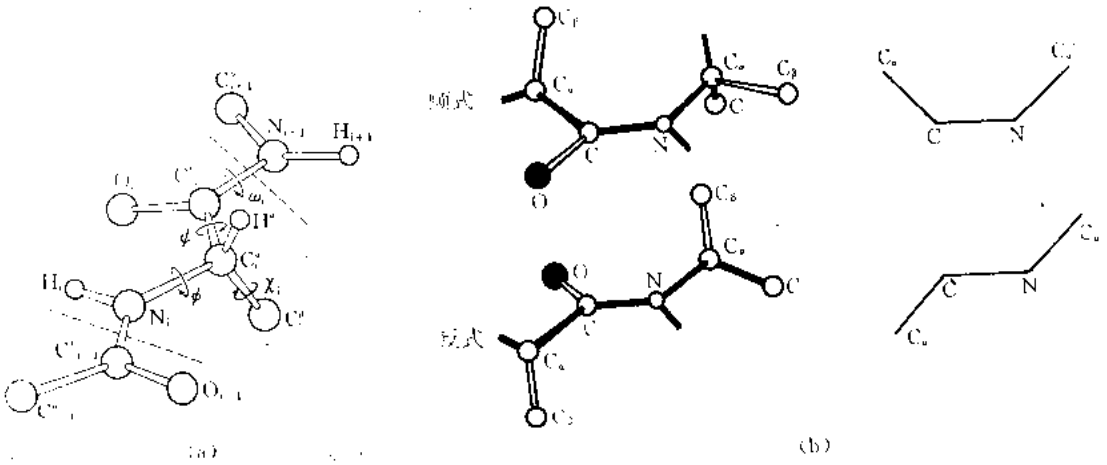


图7.6　多肽链中主链和侧链的构象角定义(a)以及肽键的顺式与反式构象(b)

这种旋转是受原子间的相互作用所限制的，即每对原子之间的接触距离不能小于两者的范德华半径之和。表7.1列出了蛋白质分子中常见原子的共价半径和范德华半径，所有原子可以看

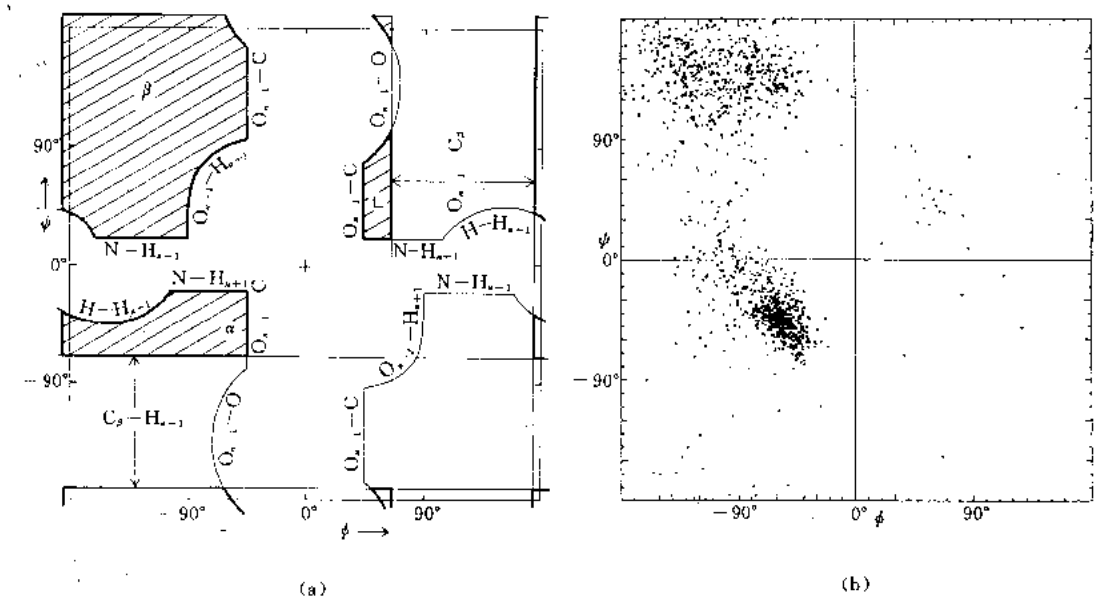


图7.7　ϕ-ψ 推导图(a)和8个蛋白质中约1000个非甘氨酸残基的实验值(b)

成是具有各自的范德华半径的坚硬的球。图7.7是根据半经验的原子间可能允许的不致于使球状原子相互碰撞的 ϕ-ψ 关系推导图。由于这种图示法最早是由 G. N. Ramachandran 等人[21,22]

提出的，因此又叫做 Ramachandran 构象图。图中标明了各对原子所对应的禁区(空白区域)，在禁区之外的区域中(斜杠区域 α、β 和 L)，ϕ 和 ψ 值的各种组合分别对应于肽链或肽链片段的各种稳定的构象(α 螺旋、β 折叠和左手 α 螺旋)。不过，甘氨酸是个例外，它允许 ϕ 和 ψ 均为0°。表7.2列出了蛋白质中一些规则构象的近似的扭角值。

表7.1 共价半径和范德华半径(nm)

元 素	共价半径			范德华半径
	单键	双键	三键	
H	0.032			0.12
C	0.077	0.067	0.060	0.20
N	0.075	0.062	0.055	0.15
O	0.073	0.060		0.14
S	0.102	0.094		0.18

表7.2 蛋白质中一些规则结构的扭角近似值

规则结构	ϕ(°)	ψ(°)	ω(°)
完全伸展的肽链	+180	+180	+180
右手 α 螺旋	−57	−47	+180
左手 α 螺旋	+57	+47	+180
平行 β 折叠片层	−119	+113	+180
反平行 β 折叠片层	−139	+135	−178

7.2 各种氨基酸对蛋白质三维结构的影响

许多蛋白质晶体结构的分析证实了早年的预言，即疏水残基存在于蛋白质分子的内部，而亲水残基多分布在与水接触的表面上。因此，在对可溶性球状蛋白质的三维结构进行讨论时，根据极性大小和在蛋白质分子中的位置将20种氨基酸进行分类较为合适。如图7.2中所示，第一类为位于外部的残基，是指通常带电荷的或极性的一些残基，包括带负电荷的 D 和 E(在 pH7)、带正电荷的 K 和 R(在 pH7)、中性的或带正电荷的 H(决定于 pH 值)以及中性的但具有极性的 N 和 Q。所有这些残基的侧链都是亲水的，通常在蛋白质分子的表面，与溶剂接触。除了在某些特殊的部位如酶的活性中心等以外，它们很少埋藏在内部；第二类为位于内部的残基，它们的侧链为不带电荷的烃，即 F，L，I，V 和 M(它像一个烃)，所有这些侧链通常都埋藏在蛋白质分子内部，而且这些不同形状和大小的侧链十分好地镶嵌堆积在一起形成疏水核；第三类为既可位于内部，又可位于外的残基，包括较小的 G 和 A、具有羟基的 S，T 和 Y、具有巯基

的C以及W和P。

不同的氨基酸残基对蛋白质的折叠和三维结构还具有明显不同的效应。最重要的是，这些不同的侧链如何与主链和其他侧链相互作用以及它们各自在不同的二级结构或三级结构类型中起什么作用。特别是G、P和C具有最特殊的几何学特点，因而应首先加以讨论。

甘氨酸是侧链仅有一个氢原子的最小的残基，也是唯一具有对称的C_α的残基。由于它没有C_β，因而具有广范围的ϕ和ψ值。它作为最小的残基，具有三种特殊的作用：在主链的两个肽段相互十分靠近的位置处常常是一个G；它可出现在其他残基的C_β靠近时所禁止的一切构象中；它比其他残基柔性大，所以它常出现于那些需要运动或转折的肽链片段之中。

脯氨酸比其他任何残基都具有更强的立体化学效应。它的环形侧链包含了主链的一部分，封闭的环使ϕ角固定为$-60°$左右，因而常常形成一个转角，改变了主链的方向。而且，由于它仅仅具有一个可变的主链构象角ψ，并缺少形成氢键所需的主链—NH—，以及它的环具有空间障碍等，所以它对规则的二级结构如α螺旋和β折叠等都具有破坏性，不过它可出现在这些二级结构的末端以及弯曲的α螺旋的中央弯曲部位。脯氨酸的另一个特殊的性质是它可以顺式肽键的形式存在(对于其他氨基酸来说，虽然顺式肽键也是可能的，但极为罕见)，因而总是形成一个转角。由于脯氨酸常出现在转角中以及很少出现在规则的二级结构中，所以它几乎总是暴露于蛋白质分子的表面。因此，虽然它的侧链是疏水性的，但是它的典型的位置是一个亲水侧链的位置。

半胱氨酸由于能结合Fe，Fe-S簇和其他一些非蛋白质基团以及能形成二硫键，因而十分特殊和重要。半胱氨酸以三种不同的类型出现，即自由的—SH基团、配位的—SH基团以及二硫键。自由的—SH型式的半胱氨酸虽然结构上像丝氨酸，但是两者的性质是十分不同的。半胱氨酸比丝氨酸形成氢键的能力差得多，而且丝氨酸是亲水的，因而是典型的暴露在分子外部的，而半胱氨酸(甚至于作为自由的—SH型式时)则是埋藏在内部的。作为配位体型式的半胱氨酸，它的—SH基团能与各种不同的金属和非蛋白质基团结合，包括Fe，Zn，Cu，Fe-S簇以及血红素等。典型的情况是，在三维空间中半胱氨酸成簇地聚集在一起，共同与一个金属或基团形成配位结合，而且一些或所有的这样的簇常常在一级序列中也是十分接近的。二硫键的形成对于稳定天然蛋白质的最后的三维结构具有最特殊的和最强的功能，即在一个蛋白质分子折叠已达到了它的最低能量时，二硫键能增加它的稳定性。

丙氨酸、缬氨酸、亮氨酸、异亮氨酸、蛋氨酸、苯丙氨酸、酪氨酸、组氨酸以及色氨酸都具有疏水性的脂肪族或芳香环侧链。这些侧链的形状的多样性，就像拼板玩具那样，在肽链骨架千变万化的各个不同的蛋白质分子内部形成了侧链镶嵌得十分好的疏水核心。其中，丙氨酸由于侧链很短，又无化学活性，因而在分子内部或表面都十分合适。酪氨酸和色氨酸具有芳香环侧链，但它们各自可形成一个氢键，因而它们可位于分子内部或外部，它们和具有完全疏水性侧链的苯丙氨酸是蛋白质内部疏水核心中的主要成分。它们的芳香环侧链，朝着骨架肽，朝着蛋氨酸或胱氨酸的脂肪族侧链或硫，或者互相对着而堆积起来。蛋白质中侧链的环与环的堆积是较为不规则的，实际上芳香环经常以相互垂直的人字形方式堆积着，而不是像在核酸中碱基对系统那样的平行的堆积。由于芳香环是大的和刚性的结构，具有相当少的构象自由度，因而它们的存在和局部的构象对周围的结构具有很大的影响。蛋氨酸中的硫虽然也可能形成氢键，但蛋氨酸的主要功能是提供一个十分柔韧的疏水侧链，因而它虽然也可存在于分子表面，但大多数是埋藏在内部的。组氨酸的最主要的功能是作为活性位置的一个残基。它的侧链咪唑环具有

一个电荷基团，它是唯一的 pK 值接近中性的特殊氨基酸，因此在其周围环境中发生的微小变化即可使它获得或者失去电荷。组氨酸的正电荷常与一个埋藏着的活性基团的负电荷相结合，而且它还能与蛋白质中的其他一些基团形成氢键。

天冬酰胺和谷酰胺的侧链末端是与肽键基团相象的酰胺基团，因而具有许多特殊的性质。天冬酰胺的酰胺基团可以以氢键与主链相结合。天冬酰胺还是除了甘氨酸以外仅有的被发现可存在于左手 α 螺旋中的氨基酸。有趣的是，谷酰胺虽然与天冬酰胺那么相象，但由于它的侧链多了一个次甲基，它不具有上述天冬酰胺的性质，它只能形成一般的氢键，也不出现在左手 α 螺旋中。

天冬氨酸、谷氨酸、赖氨酸和精氨酸的侧链在一般情况下都带有电荷，因而有两个最主要的性质：一是位于分子的外侧，尤其是赖氨酸和精氨酸，它们存在于蛋白质分子表面，仅有很少例外；二是能与相反电荷的基团形成盐键。盐键是普遍存在的，最常见的是在一级结构序列上远距离的残基之间的盐键。多肽链的 C 末端常像天冬氨酸或谷氨酸一样作为负电荷而起作用，N 末端则像赖氨酸一样作为正电荷而起作用。在蛋白质分子表面上，相邻的电荷可直接相互作用，也可不直接地或通过有序的水分子而相互作用。在蛋白质分子内部，很少有电荷，即使有也是成对的相反电荷之间形成的很好的盐键。天冬氨酸和谷氨酸在与电荷有关的功能中是十分相似的，不像天冬酰胺和谷酰胺那样性质大不相同。天冬氨酸和谷氨酸，特别是谷氨酸能与钙结合。这些位置常常在分子表面，而且在序列中紧密地形成簇。这种在分子表面的不对称的电荷分布形成了与另一个大分子相互作用的位置。赖氨酸和精氨酸除了可与天冬氨酸或谷氨酸形成盐键外，还可与带负电荷的大分子如核酸的磷酸基团形成特异性的或非特异性的盐键。这两个氨基酸都十分经常地涉及到活性位置，其中赖氨酸常与结合功能有关，而精氨酸则常与催化作用有关。尽管两者都具有电荷和长链，但在许多方面两者是不十分相同的。赖氨酸是所有侧链中最可活动的，在电子密度图中常常看不到 C_β 或 C_γ 以外的原子，而且尽管它的侧链的大部分具有相当大的疏水性，但它的侧链表面区域是被埋藏得最少的残基。精氨酸的侧链则仅仅是偶尔无序的，而且被埋藏的部分也较多，常常沿着侧链形成广泛的范德华相互作用，甚至于卷曲以形成一个疏水的表面。精氨酸的末端除了带正电荷以外，还具有5个氢键的供体存在于一个大的刚性的平面中，能形成若干个与蛋白质的氧或与有序的水分子的氧之间的氢键。

丝氨酸和苏氨酸具有短链的—OH 基团的侧链，因而它们(尤其是丝氨酸)是化学上活泼的，并因此常作为活性部位的残基而具有十分重要的功能。由于—OH 基团可作为氢键的受体或者供体，因而丝氨酸和苏氨酸常常以氢键连接到相邻骨架肽键基团—NH— 或 —C=O 或者溶剂中的有序的水分子上。不过，由于它们的侧链较短，而且只有一个可变的侧链构象角，因而形成氢键受到一定的限制。苏氨酸的—OH 基团不能直接与肽键基团形成氢键，但可通过一个水分子而间接地相互作用。

7.3 纤维状蛋白质的立体结构

用 X 射线衍射方法研究蛋白质的结构是从20世纪20年代研究纤维状蛋白质的 X 射线纤维衍射图开始的。自1931年 W. T. Astbury[233]的开拓性工作以后，研究得到了广泛和深入的进展。

这类蛋白质一般具有下列4个特性：不溶于水；在一定限度内伸长之后还可以复原；不被一般的蛋白水解酶所水解；具有不同于球状蛋白质的特殊的氨基酸组成。根据X射线纤维衍射花样的不同，可将纤维状蛋白质分成3类，即α角蛋白类(α螺旋型)，β角蛋白类(β折叠型)和胶原类(三股螺旋型)。

7.3.1 实验方法

纤维状蛋白质的结构是用X射线纤维衍射照相法进行研究的，原理和衍射花样见图7.8。

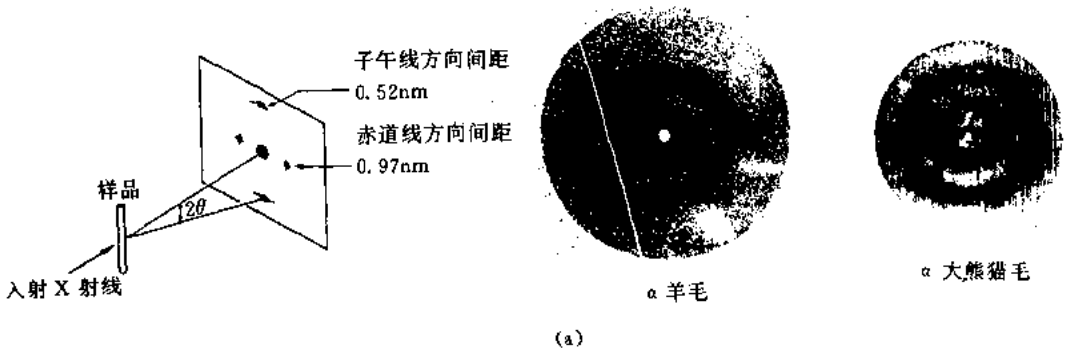
在我们实验室中，将大熊猫毛捆成直径约1.5mm的束，尽可能使束中所有纤维相互平行。将毛束固定在平板照相机上，用CuK_α-Ni滤波的X射线(40kV，100mA)经过直径为0.8mm的准直器垂直于纤维轴射入。样品与底片的距离为10cm，照射3小时，结果见图7.8(a)。

衍射效应是许多单根纤维或亚纤维成分的效应的总和。假若每根纤维的结构差别不是很大的话，那么得到的衍射花样基本上与一个单根纤维的衍射花样相同。不过，由于纤维中内部结构的变形，每根纤维不是理想的三维单晶，而且粗大的纤维束样品中各个纤维不是很好地平行排列的，因此，通常在单晶衍射图中分立的各个衍射点都变成为一个等半径的圆弧(如果纤维束中各纤维的方向是完全随机的话，那么每个衍射点就会变成一个整圆)，而且是弥散的。纤维衍射图中一般只有少量的衍射斑点，通常分布在代表了两种间距的衍射范围内：一是广角衍射，衍射点反映了小的间距，代表了原子、原子团、氨基酸残基以及分子的一部分的大小(2nm以下)；第二个范围是小角衍射(可小至4°)，衍射点反映了较大的间距，代表了亚纤维等结构的大小(2—100nm)，与电子显微镜的测量范围重叠。纤维图中子午线上的衍射可以用来测定纤维结构的轴向周期，赤道线上的衍射用来测定纤维结构中的横向周期。可根据布拉格方程进行计算：$2d\sin\theta=n\lambda$，$n=1$通常是最强的反射，由实验测出某个衍射点的θ值，即可由$2d\sin\theta=\lambda$计算出这个衍射斑点所代表的结构的最小重复距离d。对于较小的θ来说，$\sin\theta\approx\theta$，因此可简化为$d\approx1/\theta$。

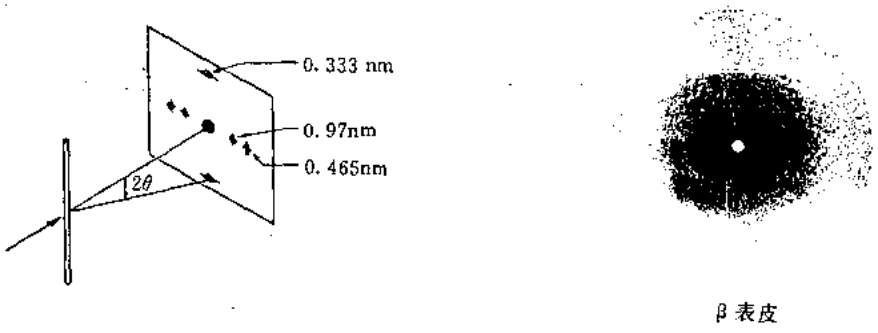
研究和解释蛋白质的纤维衍射图主要靠建造模型的方法和理论计算。值得提及的是，早期研究毛发、蹄角等角蛋白的X射线纤维衍射花样时，人们设计了许多螺旋模型来解释衍射斑点，但这些模型都跳不出整数的概念，即螺旋每上升一圈所包含的残基数为整数。用这些模型推算出的衍射斑点都与蛋白质纤维衍射图的花样不符。1951年L. Pauling和R. Corey[24,25]根据氨基酸和小肽的键长和键角等立体化学实验数据，对多肽链的构象提出了两个原则，称为鲍林规则：第一，肽键酰胺基团应具有平面构象；第二，多肽链的构象应满足结合最多限度氢键，主链氢键—NH…O为0.28nm左右，而且N→H与N→O两个向量的夹角应不大于30°。他们用精确的模型法试验了各种可能的多肽链构象，预言了两种具有周期性的多肽链结构，分别称为α螺旋(α-helix)和β折叠片层(β-pleated sheet)。他们提出的螺旋模型冲破了整数的观念，指出螺旋每圈包含3.6个残基，很好地解释了纤维衍射图上的斑点。按照这个螺旋模型，每个残基沿轴向应上升0.15nm。但是人们过去一直未注意到相应于周期为0.15nm的衍射点，因而当时尚有人怀疑这个模型的正确性。后来，人们在箭猪毛等的纤维衍射图中找到了这个衍射点，确证了α螺旋模型的正确性。L. C. Pauling由于揭示了化学键的本质以及确定了肽键和蛋白质分子的α螺旋二级结构等成就而获得了1954年诺贝尔化学奖。

7.3.2 α角蛋白类

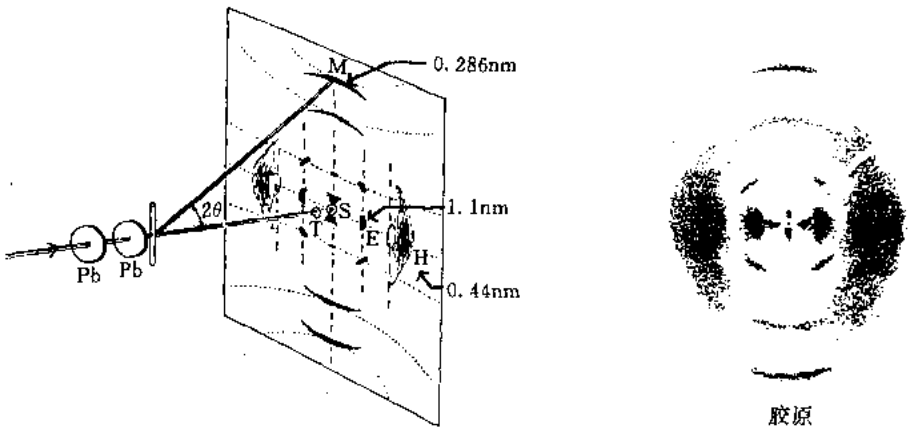
这类蛋白质包括α角蛋白(α-keratin)、肌球蛋白(myosin)、表皮(epidermis)和血纤维蛋白原(fibrinogen)等，因而又称为K-M-E-F类。α角蛋白存在于动物的毛发、角、爪、蹄、羽毛、甲



(a)



(b)



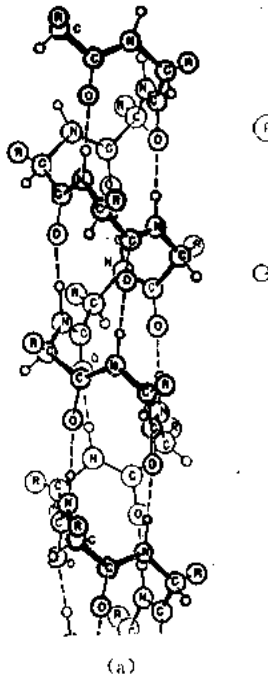
(c)

图7.8 α角蛋白(a)、β角蛋白(b)和胶原(c)的纤维衍射图

壳、指甲以及表皮中。羊毛纤维是典型的角蛋白，其中疏水性氨基酸和半胱氨酸的含量较高。

在这类纤维状蛋白质的X射线纤维衍射花样[见图7.8(a)]中，赤道线上有一对相当于纤维中横向间距为0.97nm的强度弥散的衍射斑点，同时在子午线上有一对相当于纤维轴向间距为0.52nm的弥散的弧线，在一些衍射图上还可找到一对相当于每一个氨基酸残基在螺旋轴上的投影为0.15nm的反射。Pauling和Corey指出这类纤维蛋白的基本结构单位是右手α螺旋，见图7.9示意图。在右手α螺旋模型中，螺旋每上升一圈，包含3.6个残基，因而侧链以大

约100°的方位角向外伸出；螺距为0.55nm，因而每个残基沿螺旋轴方向上升0.152nm；螺旋的直径为1nm；主链中每个残基(n)的—NH都与其后的第4个残基(即$n+4$)的—C=O形成—NH···O=C—氢键。氢键的键长为0.28nm，它的N，H和O三个原子位于一直线上或偏离不远，并与螺旋轴接近于平行，每个氢键与主链所形成的环中包含13个原子。右手α螺旋模型能很好地满足肽键酰胺基团的平面性和满足形成最多限度氢键的原则，在立体化学上也很合理。根据这个模型计算的衍射点的位置和相对强度与实验所得的衍射花样相符。不过，还必须修正，因为α螺旋的周期应是0.55nm，而衍射花样所表明的实际周期一般为0.52nm；而且若将α螺旋按照最密方式堆积时，理论密度仍较羊毛的密度低等等。为了解释这些事实，Pauling和Corey指出，在α角蛋白中，α螺旋的轴不是直线，其本身也是一个周期很长的螺旋线，这样的修正并不会显著地影响α螺旋本身的立体化学特点。现在人们一般接受的α角蛋白的结构模型见图7.10：3个右手α螺旋通过许多二硫键互相交联缠绕，形成一个直径约2nm的缆状的左手超螺旋，称为亚原纤维(subfibril)；11个亚原纤维再以9+2的结构形成更复杂的原纤维(fibril)，即9个亚原纤维围绕着另外2个亚原纤维形成直径约为8nm的原纤维；几百个原纤维结合



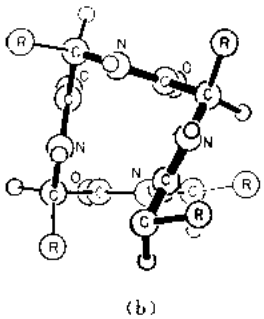


图7.9 α螺旋模型
(a)侧面观；(b)顶面观

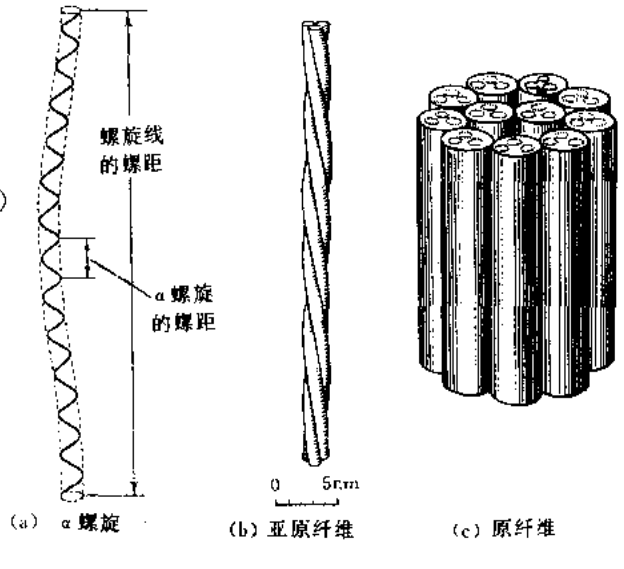


图7.10 α角蛋白模型

成直径约为200nm的粗纤维(macrofibril)。许多粗纤维平行成束，组成细胞；这些细胞平行堆积，形成毛发等。这些纤维在一定条件下可以被拉长并且具有弹性，这是由于氢键被破坏而导致α螺旋转变成了平行的β折叠结构。拉长了的纤维的X射线纤维衍射花样与β角蛋白的相同。

7.3.3 β角蛋白类

如前所述，毛发等α角蛋白如果被拉长至原来长度的两倍就变成了β角蛋白。α角蛋白与β角蛋白可以互相转化，两者的X射线衍射花样大不相同。在β角蛋白的衍射图中[见图7.8(b)]，赤道线上有两对分别相当于纤维中横向间距为0.97nm和0.47nm的强度弥散的斑点，在子午线上有相当于纤维轴方向间距约为0.333nm的斑点。β角蛋白类的另一个典型例子是天然蚕丝的丝心蛋白，它的X射线衍射花样表明，沿纤维轴的周期是0.7nm。Pauling和Corey提出了β折叠片层的结构模型(见图7.11)来解释所有这些事实。他们指出，一个完全伸展的多肽链应是主链共面的完全伸展的链，如图7.4中所示，它的周期应为0.727nm。而天然丝和β角蛋白纤维的周期各为0.7nm和0.666nm(0.333nm×2)，因而，在天然丝中，多肽链几乎是完全伸展的；而在β角蛋白中，除了酰胺基团仍维持共面外，多肽链中的C。相连的两个键可能发生旋转，因而已不是一个共面的完全伸展的链了，所以链的周期缩短了。天然丝和β角蛋白分别属于反平行β折叠片层结构和平行β折叠片层结构。在这些结构模型中，两条以上的多肽链或一条多肽链的不同片段相互平行，并由链间氢键结合形成折叠的片层，侧链在片层的上、下方交替伸出。如果相互平行的肽链从N末端到C末端方向相同，称为平行β折叠；方向相反则称为反平行β折叠。在平行β折叠中，链内周期为0.65nm，链的垂直方向的周期即链间距离是0.47nm。主链上每个—NH基团和—C=O基团都能形成链间的氢键—NH···O=C—。氢键不垂直于肽链，氢键使肽链间形成相同宽度的空间。在反平行β折叠中，肽链几乎完全伸展。链内周期为0.7nm，此时链的垂直方向的周期等于链间距离的两倍，即大约1nm。氢键与链垂直，并使肽链形成宽窄相间的空间。这两个模型都能满足肽键酰胺基团的平面性原则和形成最多限度氢键的原则，并能很好地解释衍射花样。

7.3.4 胶原类

胶原广泛地存在于动物体中，占机体总蛋白质的25—30%，是皮肤、骨胶、软骨、筋、动脉管壁以及结缔组织中的蛋白质成分。它们的氨基酸组成十分特殊：Gly(33%)、Pro(12%)以及Hypro(9%)的含量特别高。胶原类的X射线纤维衍射图见图7.8(c)，在赤道线上有两对各相当于间距为0.44nm和1.15nm的强衍射斑点；在子午线上，有较强的一对圆弧形衍射斑点与纤维的轴向周期0.286nm相对应。胶原的结构不同于α螺旋和β折叠。实际上它不可能形成α螺旋，因为肽链中Pro和Hypro的含量太高，它们是α螺旋的破坏者。胶原的分子模型见图7.12，它是直径约1.4nm，长约280nm的棒状分子，由三个多肽链组成。每个多肽链是一个略微向左扭曲的左手螺旋，三个多肽链再相互盘绕成一个大的右手螺旋，称为三股螺旋(triple helix)或超螺旋(super helix)。每个残基在这个超螺旋中上升0.29nm，每圈的残基数约3.3。三个肽链以氢键互相结合，氢的供体是Gly的肽键—NH基团，氢的受体是在另一个肽链上的肽键—C=O基团。各条链上每第三个残基都是Gly，因为三股螺旋的内侧是很拥挤的，只有Gly才能位于内部并与其他两条链形成氢键，其他残基Pro和Hypro的大的环形侧链都在三股螺旋的外侧。由于胶原类结构的复杂性，这些蛋白质的纤维图上还呈现出相当于8.5nm和64nm的超周期的衍射弧等。

7.4 球状蛋白质的晶体结构

生物体中最重要的结构蛋白质和功能蛋白质中的大多数是球状蛋白质，它们遍布于生物

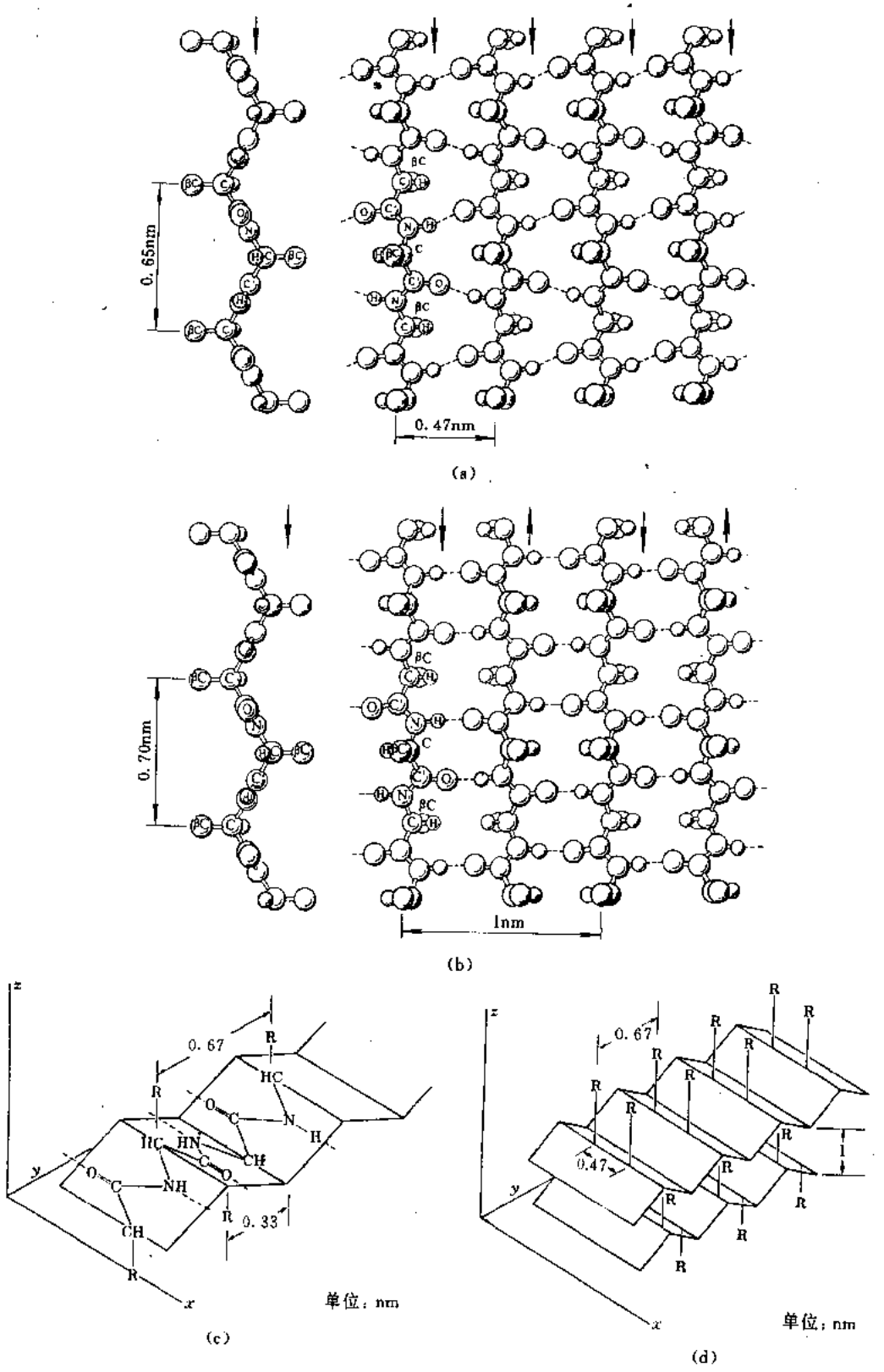


图7.11 β折叠片层示意图

(a)平行β折叠片层；(b)反平行β折叠片层；(a)和(b)中左侧为片层的侧面观，右侧为片层的正面观。(c)和(d)为伸展肽链的折叠结构和折叠片层示意图

体中。球状蛋白质的空间结构比纤维状蛋白质复杂得多，它们不是简单地沿着一个轴有规则地重复排列，而是在三维空间中卷曲、折叠、盘绕而成紧密的近似于球形的结构，即如前所述，可以被描述为初级结构（即一级结构）和高级结构（包括二级结构、三级结构及四级结构）。高级结构的测定对于阐明结构与功能的关系具有重要意义。

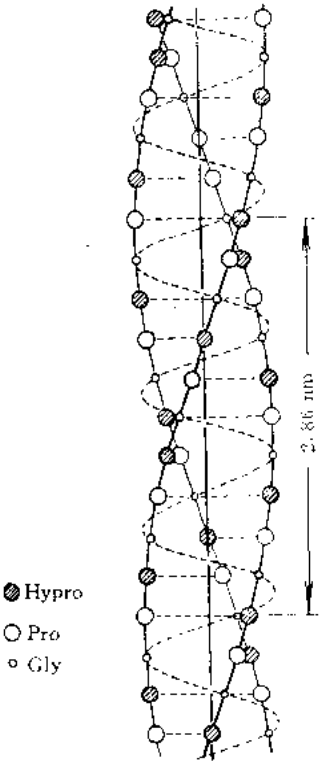


图7.12　胶原的结构模型

7.4.1　实验方法

测定球状蛋白质的晶体结构采用X射线单晶衍射法。首先必须生长出适合于X射线衍射用的较大的蛋白质单晶（一般线度达0.5—2mm）。蛋白质晶体是“湿晶体”，即在蛋白质分子之间充满着母液，仅仅当晶体在与这种液体或它的蒸汽平衡时才是稳定的。X射线衍射方法所得到的高分辨率电子密度图表明，一些水分子结合在蛋白质分子表面上，形成了有序的水分子“外套”；在内部，有序的水分子与蛋白质形成氢键网。当晶体干了，即失去了结晶水时，晶胞收缩，晶体就变成了无序的固体。晶体质量的好坏直接决定了衍射的分辨率和结构测定的精确度。我们知道，早在1926年，J. B. Sumner[26]报道了第一个被结晶的蛋白质即脲酶晶体，随后J. H. Northrop以及W. M. Stantey又分别获得了胃蛋白酶、过氧化氢酶、过氧化物酶、脂肪氧化酶、胰蛋白酶、胰凝乳蛋白酶、白喉毒素、烟草花叶病毒等的晶体。证明了酶是一种蛋白质，而且可以结晶，因而揭示了酶和蛋白质这一重要生命物质的化学本质，他们三人因此而获得1946年诺贝尔化学奖。胃蛋白酶是第二个被结晶的蛋白质，它的晶体是第一个可用于X射线衍射的球状蛋白质晶体（2mm长），并于1934年由J. D. Bernal[27]拍摄了第一张蛋白质单晶的X射线衍射照片。不过，当时未能解出它的晶体结构。早年每一个新的蛋白质晶体和衍射照片都会令人十分激动和鼓舞，它表明了这些大分子与化学小分子一样在晶格中是有序的，因而可以用X射线衍射法来测定它们的结构。到现在，已有许多球状蛋白质被结晶了，而且作为一个规律，只要球状蛋白质是很纯的，就可能在体外形成晶体。不过需要人们去试验，找到合适的晶体生长条件。在我们的实验室中，也进行了多种蛋白质的晶体生长试验，得到了可适用于X射线衍射的蛋白质单晶，见图7.13。其中如斑头雁高铁血红蛋白的晶体生长方法和条件为[28]：用悬滴式微量汽相扩散法。悬滴中含5μl血红蛋白溶液（$A_{280}=0.43$）、2μl20%PEG6000和3μl0.05mol/l磷酸钾缓冲液，pH7.0；池中平衡液为1ml 12% PEG6000。置23--25℃恒温箱中，一星期后可生长成合用的单晶。其X射线衍射旋进照片见图7.14。

与用X射线纤维衍射照相法研究纤维状蛋白质不同，球状蛋白质的单晶的X射线衍射分析方法是：首先用旋进照相法测定晶胞参数和空间群，然后用四圆衍射仪、平面探测仪或其他各种照相方法记录各衍射点的强度。并用多对同晶置换法或将反常散射法与其相结合，测定出未知结构的各个衍射点的相角；或者利用同源蛋白质或类似蛋白质的已知晶体结构通过分子置换法（包括旋转函数法和平移函数法）来确定未知晶体的各个衍射点的相角。最后计算出未知结构的电子密度图，建立模型并进行修正。

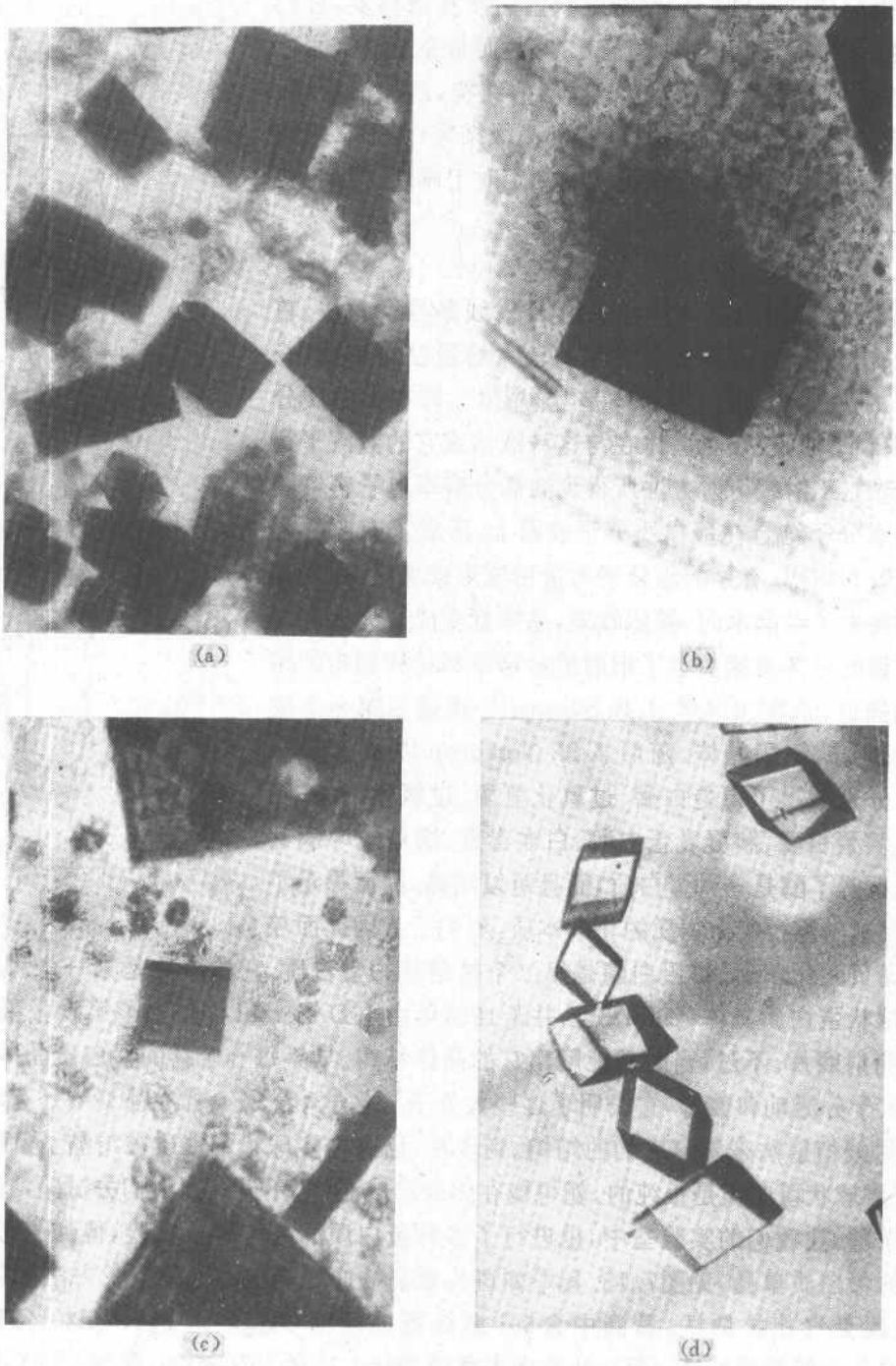


图7.13　蛋白质晶体

(a)斑头雁高铁血红蛋白[28]；(b)斑头雁氧合血红蛋白[29]；(c)胰蛋白酶自溶活性产物——绿豆胰蛋白酶抑制剂复合物[30]；(d)抗菌多肽 LC1[31]

世界上第一个被测出晶体结构的球状蛋白质是肌红蛋白。J. C. Kendrew 等[32]于1960年发表的抹香鲸肌红蛋白的0.6nm 分辨率结构模型，首次揭示了多肽链的排列，见图7.15(c)。后来基于0.14nm 分辨率的结构分析，最终测出所有原子的排列。他们根据 F. A. Perutz 等人[33]关

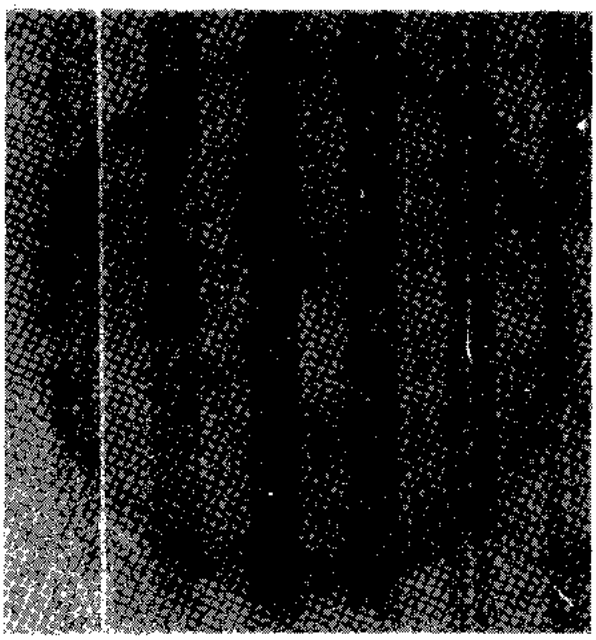


图7.14 斑头雁高铁血红蛋白晶体的旋进照片

CuK_α，单色器滤波，40kV，230mA，晶体与底片距离为65mm，曝光30小时。

于同晶置换法可以用于解决蛋白质的相角问题的证据，首次成功地将同晶置换法应用于蛋白质晶体结构测定，解决了困惑多年的相角问题。J. C. Kendrew 和 F. A. Perutz 两人因此而获得1962年诺贝尔化学奖。Kendrew 采用共同结晶的方法[34]制备了5个合用的重原子衍生物，即将重原子试剂如对-氯汞苯磺酸、二氨汞以及氯化金等重原子试剂分别与肌红蛋白同时配制于缓冲母液中生长晶体，使重原子结合到蛋白质的特殊基团或位点上。一般，制备重原子衍生物用浸泡法，因为共同结晶法容易使蛋白质变性等。浸泡法是将重原子试剂配制于蛋白质晶体生长的缓冲母液中，然后将蛋白质晶体浸泡于其中，重原子试剂通过扩散而结合到蛋白质的特殊基团或位点上。以笔者在对大肠杆菌的麦芽糖结合蛋白进行晶体结构分析的工作中制备重原子衍生物为例[35]，将重原子试剂以0.5 5m mol/l 浓度的量加至麦芽糖结合蛋白晶体生长用的柠檬酸缓冲液(pH6.2，不含麦芽糖结合蛋白)中，充分搅拌后，核对或重新将 pH 调回至6.2。然后将麦芽糖结合蛋白晶体浸泡于其中一定时间以后，取出晶体，用旋进照相法观察衍射点强度是否有变化，或者用四圆衍射仪收集分辨率为2- 0.6(或0.5)nm 的一个独立区的衍射数据，直接计算差值帕特逊函数来确定是否有重原子以及它们在晶胞中的坐标。为了防止某些重原子试剂在柠檬酸缓冲液(或磷酸缓冲液)中生成沉淀，可以用 pH 值相同的醋酸缓冲液等代替柠檬酸缓冲液(或磷酸缓冲液)。在试验了分别含有 Ag，Au，Cd，I，Ir，Hg，Os，Pb，Pt，Re，Sm，Ti，U，W 等14种重原子的共36种试剂后，得到了9个同晶型重原子衍生物，其中有5个较好，用于多对同晶置换法进行结构分析，例如 $Pb(NO_3)_2$，1m mol/l，浸泡12天；$K_2Pt(NO_3)_4$，1m mol/l，浸泡8天；以及 mersayl-Hg，1m mol/l，浸泡13天；此外，还利用该蛋白质与麦芽糖结合的特性，用碘化物6-I 麦芽糖，10m mol/l 浸泡22天，也得到了合用的重原子衍生物。

自从肌红蛋白解决了相角问题并测定了晶体结构以后，相继在1965年 C. C. F. Blake 等[36]测定出溶菌酶、1967年 G. Kartha 等[37]以及 H. W. Wyckoff 等[38]测定出核糖核酸酶、B. W.

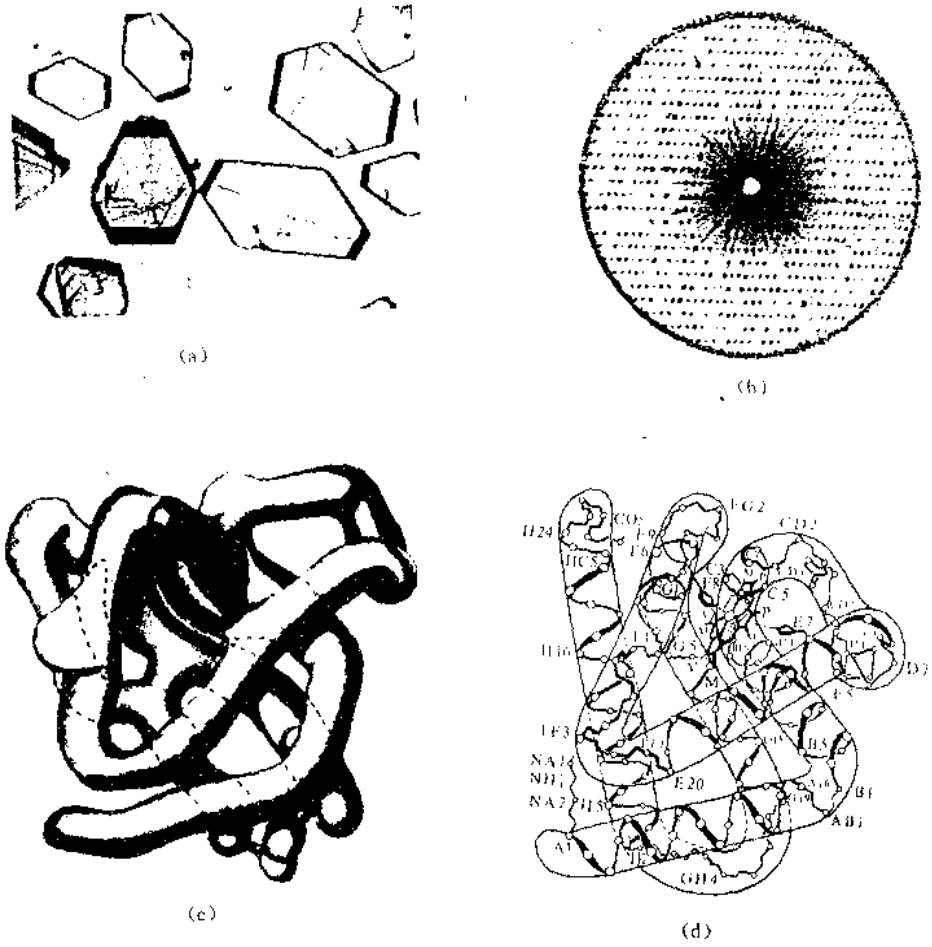


图7.15　抹香鲸肌红蛋白的晶体(a)，旋进照片(b)，0.6nm分辨率模型(c)以及0.14nm高分辨率结构模型(d)

Metthews等[39]测定出胰凝乳蛋白酶以及G.N.Reeke等[40]测定出羧肽酶的晶体结构等。由于实验技术和计算机技术的迅速发展，被测定了高分辨率晶体结构的蛋白质数目增加很快，现在已达3千多个。我们不能逐一介绍，但其中值得提及的一个是J.Deisenhofer等人[41]于1984年发表的绿色红假单胞菌的光合作用中心的晶体结构，这是第一个膜蛋白的分子结构，并具有重大生物学意义，将在后面的章节中专门讨论。

在我国，由中国科学院和北京大学(包括本实验室在内)合作组成的胰岛素结构研究组[42]于1972年报道了0.25nm分辨率猪胰岛素的晶体结构，这是我国第一个蛋白质的晶体结构分析工作，是用浸泡法制备重原子衍生物，并用多对同晶置换法分析出晶体结构。

现在，人们已有可能从大量晶体结构资料中总结出关于球状蛋白质的二级结构、三级结构和四级结构的一些规律性，下面将分别进行介绍。[10,43,44,45]

7.4.2　球状蛋白质的二级结构单元

在第一个蛋白质单晶的晶体结构被测定以前，人们所想象的蛋白质分子的空间结构是多

肽链的相当简单的几何排列，例如笼形、重复的之字形或均匀排列的平行棒等。确实，最初Pauling等人推测出来的纤维状蛋白质的结构就是这样的，即典型的α螺旋或β折叠重复结构。抹香鲸肌红蛋白的0.6nm分辨率模型以及1961年进行的高分辨率分析，最后结论性地证实了α螺旋模型的正确性，并证明了α螺旋也存在于球状蛋白质中，而且确实是右手α螺旋。后来大量的蛋白质单晶三维结构被测定，表明了α螺旋和β折叠结构也普遍存在于球状蛋白质中，而且它们以各种方式结合在一起。但是，球状蛋白质的结构复杂得多，除了α螺旋和β折叠以外，还存在其他一些二级结构类型，下面将分别描述它们。

7.4.2.1　α螺旋、3_{10}螺旋和π螺旋(α-helix、3_{10}-helix、π-helix)

如果多肽链中每个残基具有相同的φ角和ψ角，则会形成螺旋。但是能满足所有构象要求的螺旋结构却是有限的。这些构象的要求是：在允许范围内的扭角φ和ψ值、合适的范德华接触以及形成尽可能多的距离合适的近乎直线的氢键等等。由于氨基酸具有L构型，限制了只能是右手螺旋(左手螺旋中羰基氧与C_β靠得太近。不过，在嗜热菌蛋白酶中发现存在有左手螺旋)。而且，从理论上讲，可能存在的右手螺旋只有三种，可以用每圈螺旋所包含的残基数目n和每个氢键与主链形成的环中所包含的原子数目s来表示和区别它们(n_S)，即3_{10}螺旋、3.6_{13}螺旋(即α螺旋)以及4.4_{16}螺旋(即π螺旋)，示于图7.16中。

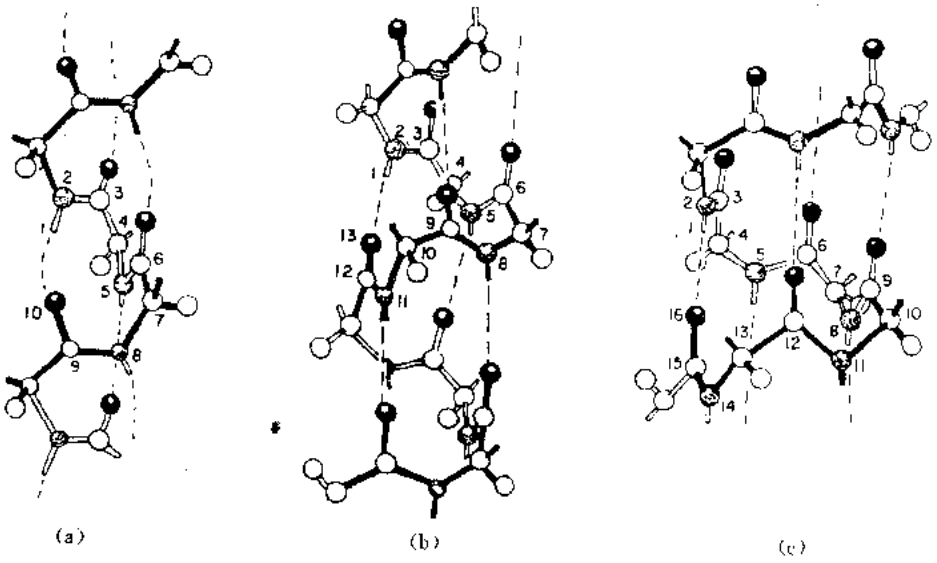


图7.16　α螺旋(a)、3_{10}螺旋(b)和π螺旋(c)的模型

右手3_{10}螺旋(简称3_{10})为螺旋每上升一圈，包含3个残基；螺距为0.6nm，每个残基上升0.2nm；主链上第n个残基的—NH—与第$n+3$个残基的—C=O形成氢键，氢键长0.28nm；每个氢键与主链所形成的环中包括10个原子。

右手3.6_{13}螺旋即α螺旋(简称α_R)，螺旋每上升一圈，包含3.6个残基，因而侧链以大约每隔100°的方位角伸向螺旋外侧；螺距为0.55nm，每个残基上升0.152nm；螺旋的直径为1nm；主链上残基n的—C=O与残基$n+4$的—NH形成氢键，氢键键长为0.28nm，氢键的N，H和O三个原子处在一直线上或偏离不远，并接近于与螺旋轴平行；每个氢键与主链所形成的环中有13个原子。

右手4.4_{16}螺旋即π螺旋(简称π),螺旋每圈包含4.4个残基,每个氢键与主链所形成的环中包括16个原子。

虽然以上三种螺旋在理论上都是允许的,但其中只有α螺旋毫无疑问是完全合理的。至今所得到的X射线衍射分析结果也表明,无论在纤维状蛋白质或是在球状蛋白质中都存在有α螺旋的构象。3_{10}螺旋的短片段也相当经常地出现,不过很少发现长的3_{10}螺旋。π螺旋则很少见到。

在球状蛋白质中,常出现最典型的规则的α螺旋(又称为α_{10}),如图7.17所示,但也发现有一些非典型的α螺旋,如图7.18中所示的α_{11}螺旋,在这种α螺旋中,氢键的N、H和O三个原子不成直线,N--H…O夹角为109°,氢键长0.29nm。从能量上看,这种螺旋与典型的α螺旋相差不远,但要继续形成第二圈螺旋就不可能了,因此它常常出现在典型的α螺旋的末端。此外,在蛋白质中还发现弯曲的α螺旋和极为特殊的受干扰的α螺旋等,见图7.18。

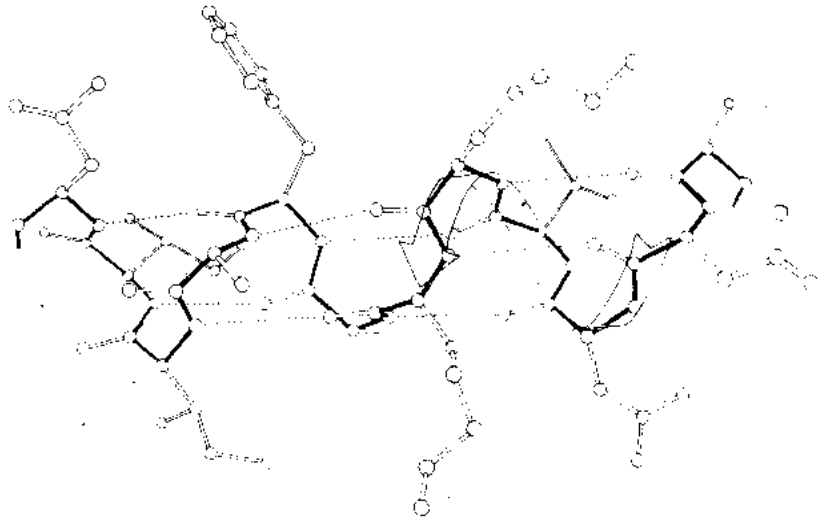


图7.17 典型的α螺旋(α_{10})实例
鲤鱼肌肉钙结合蛋白残基40—51

α螺旋是蛋白质中典型的二级结构单元。一个α螺旋能包含的残基数目可达35个之多,而最长的β折叠肽链仅包含15个残基。α螺旋比任何其他二级结构单元对蛋白质的稳定性和结构都具有更大的影响。在蛋白质分子中,α螺旋的含量有的可高达80%,有的不含α螺旋,但为数不多。

不同残基在α螺旋中出现的几率只具有十分弱的、但又是确定的倾向性。Ala,Glu,Leu以及Met是最常出现的;Pro,Gly和Ser则不常出现。另外,由于α螺旋的侧链以大约每100°的方位伸向外部,而α螺旋又最常见于蛋白质分子的外部,因此,其肽链中每隔3--4个残基,疏水和亲水侧链应交替地出现。这种倾向有时在氨基酸的序列中可以看到。不过,还不足以由此而预测α螺旋的结构。

7.4.2.2 β折叠(β-sheet)

β折叠也是由Pauling等首先从理论上推测出来,并在纤维状蛋白质和球状蛋白质中都发现确实存在的主要二级结构类型之一,见图7.11。从理论上看,平行和反平行的β折叠都是允

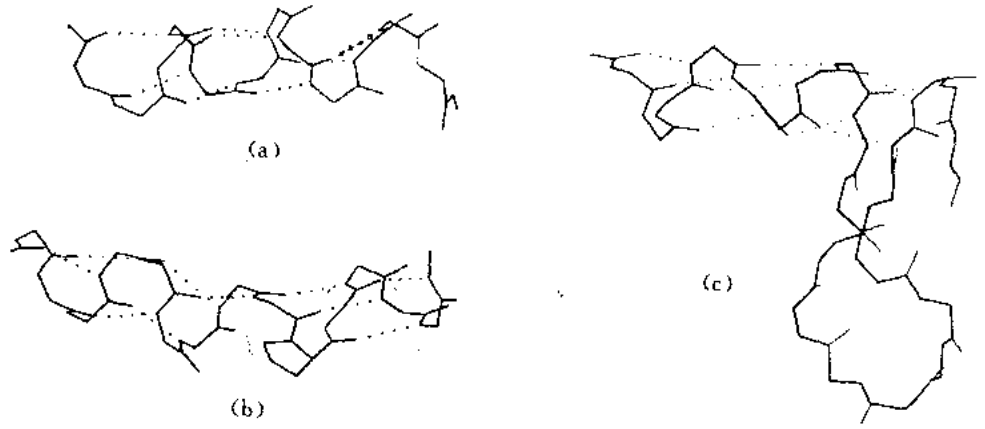


图7.18　非典型的α螺旋(α_{11})(a)以及弯曲的(b)和受干扰的(c)α螺旋实例

许存在的。不过从能量上看，反平行β折叠更为稳定。图7.19和图7.20分别为在球状蛋白质中存在的典型的平行β折叠和反平行β折叠的实例。除了一般只形成单纯的平行的或反平行的β折叠片层外，有时还看到同时具有平行和反平行链的混合β折叠片层，但是事实上存在着一种阻止形成混合片层的倾向，这可能是由于两类氢键对肽链的走向要求稍有不同所致。

平行的β折叠片层一般不少于5股肽链，而在反平行的片层中，常可看到仅有两股肽链形成的扭曲的带，见图7.21。两股肽链都扭曲很大，每5个残基约扭曲180°。

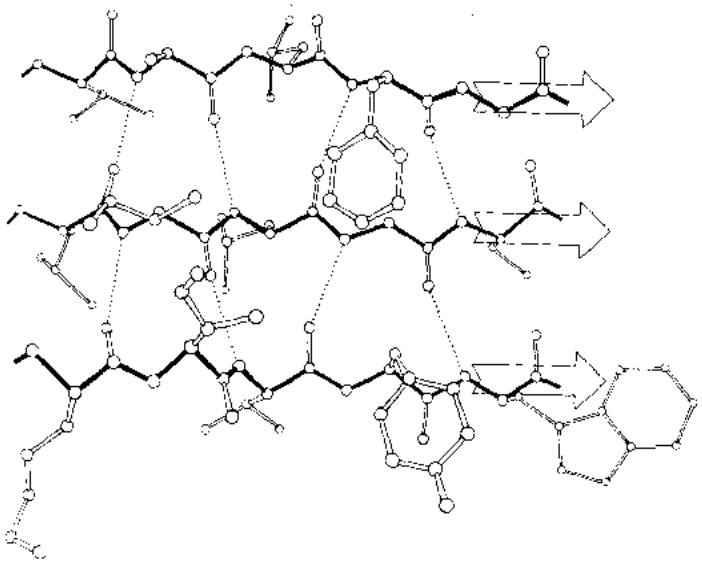


图7.19　典型的平行β折叠片层实例
黄素氧还蛋白残基82—86、49—53以及2—6

氨基酸的L手性使得β折叠片层的最显著特点之一是扭曲，而且这种扭曲总是具有右手的手性。这种情形对于实际观察到的所有球状蛋白质都是共同的。这里有三层含义：第一，伸展

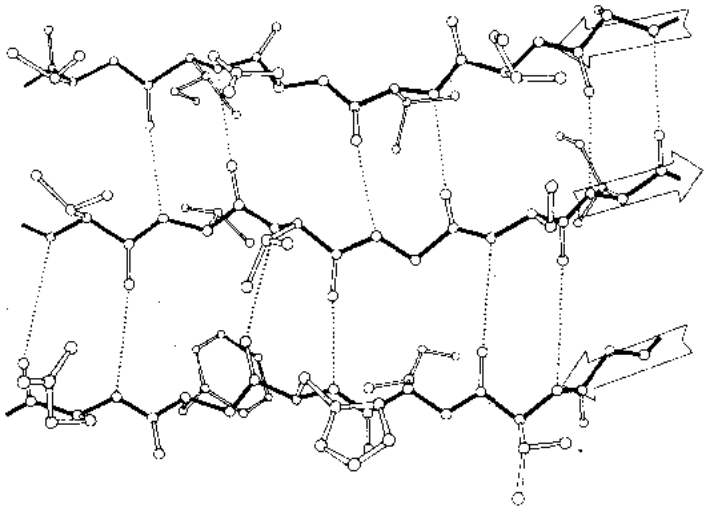


图7.20　典型的反平行β折叠片层实例
Cu,Zn过氧化物歧化酶残基93—98,28—33以及16—21

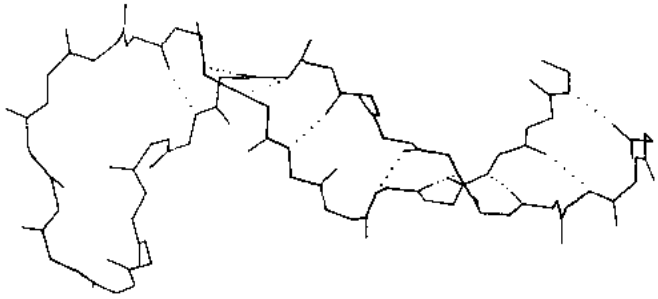


图7.21　仅具有两股肽链的带状反平行β折叠实例
乳酸脱氢酶残基263—294

的肽链本身倾向于右手方向扭曲，即若沿着多肽链轴的方向来观察伸展肽链的酰胺平面的方向或氢键的方向时，这条肽链本身的扭曲总是右手的，见图7.22。一般每个残基扭曲0—30°，通常在仅有两股链的反平行β折叠中的链扭曲角度最大，见图7.21；多股的片层中的各条链扭曲度较小。第二，由此而产生的第一个结果是，平行β股之间的交叉连接都是右手的。β折叠片层中的各链末端之间的连接方式有两种。在反平行β折叠中，链间连接为发夹型，在片层的同一端，见图7.23(a)；在平行的β片层中，则必然是交叉型连接，即从片层的一端到另一端，见图7.23(b),(c)。这种交叉型连接在理论上可以是右手的，也可以是左手的。但在蛋白质结构中实际观察到的，不论是否是连接相邻的β股，也不论连接片段的长度和构象如何，交叉连接都是右手的，只有极少数是左手的。另外，有一半以上的交叉连接中至少有一个α螺旋，见图7.24(b)。第三，为了减小构象能，伸展的多肽链的右手扭曲倾向所导致的第二个结果是，当沿着多肽链方向观看时，整个平行β片层总是再以一个右手的型式扭曲着，形成一个缓慢的大的扭曲

片层，称为马鞍形(saddle shape)或卷曲成筒形，称为β桶(β-barrels)。图7.25是这种扭曲的示意图，图7.26为实例。反平行的β折叠片层有的是不扭曲的，有的是部分扭曲的或形成β桶，见图7.27。

β桶可由5—13股肽链组成。桶的内部都是疏水性侧链。不论桶由几股链组成，桶的顶面观都呈微扁平圆形，见图7.28中的实例。平行的β片层和混合片层中的平行部分总是埋藏在蛋白质分子内部的，其他二级结构类型的主链(常常是α螺旋)在两侧保护它们；而典型的反平行β片层则为一侧暴露于溶剂，另一侧埋藏在内部，所以，它们的氨基酸序列常常是疏水性与亲水性交替出现。已知的蛋白质晶体结构资料统计结果表明，Val 和 Ile 在β折叠片层中较常出现。

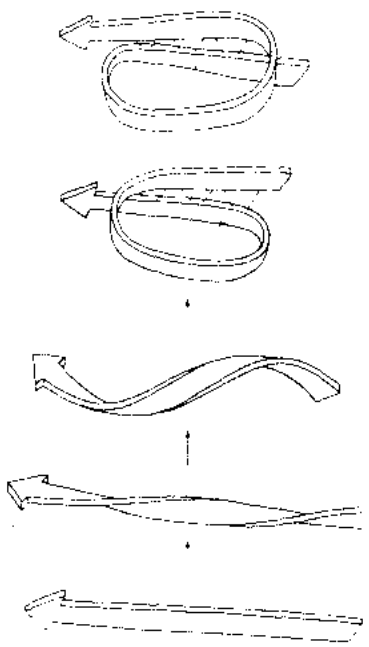


图7.22 多肽链右手扭曲的天然倾向及由此而可能形成的左手或右手连接

7.4.2.3 β转角(β-turn 或称为 β-bend 等)

球状蛋白质分子为紧密的球形，这是由于它们的肽链常常逆转方向，甚至出现180°的回折。理论分析和大量蛋白质三维结构的资料揭示了肽链的逆转主要是由一种被称为β转角的二级结构来实现的。

前面所讨论过的典型的α螺旋和β折叠是蛋白质中

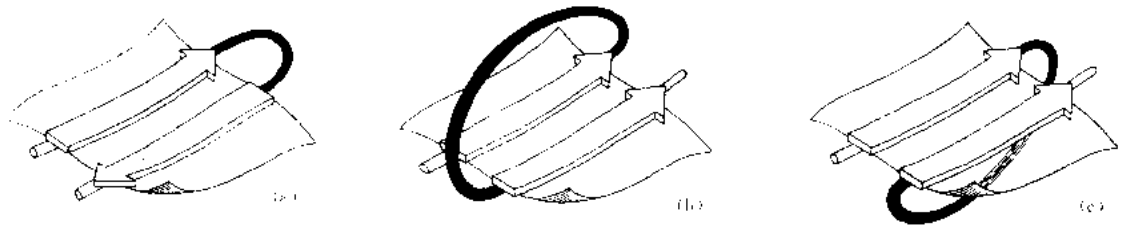


图7.23 β链之间的两种主要连接方式

(a)反平行链间的发夹型连接；(b)和(c)平行链间的右手和左手交叉连接

"规则的"重复结构，它们具有重复的主链扭角，而且它们的主链上的—N—H 和—C=O 基团之间形成周期排列的氢键。而β转角则不同，它具有非周期的主链扭角，而且典型的β转角只有一个内部的主链—NH 与—C=O 基团之间的氢键，因而是一种"非规则的"非重复结构。

β转角与α螺旋和β折叠一样，也是首先通过理论的构象分析而被认识的。从分析什么样的构象能使残基 n 的主链 C=O 与残基 $n+3$ 的主链 NH 生成氢键而形成4个连续残基的稳定结构表明，有三种可能的类型，见图7.29。类型Ⅰ的各残基的构象角为 $\phi_2=-60°$，$\psi_2=-30°$；$\phi_3=+90°$，$\psi_3=0°$。它的4个 C_α 是不共面的，而是形成大约+45°的双面角。类型Ⅰ是在球状蛋白质中确实最普遍地存在的转角，因而又被称为"普通"转角。类型Ⅱ则为 $\phi_2=-60°$，$\psi_2=+120°$；$\phi_3=+90°$，$\psi_3=0°$，它的4个 C_α 和氢键近似于在一个平面上。类型Ⅱ是确实存在的第二种主要的转角类型，其中大约60%的 R_2 是 Gly，因而又被称为"甘氨酸"转角。类型Ⅰ和Ⅱ不是螺旋，而

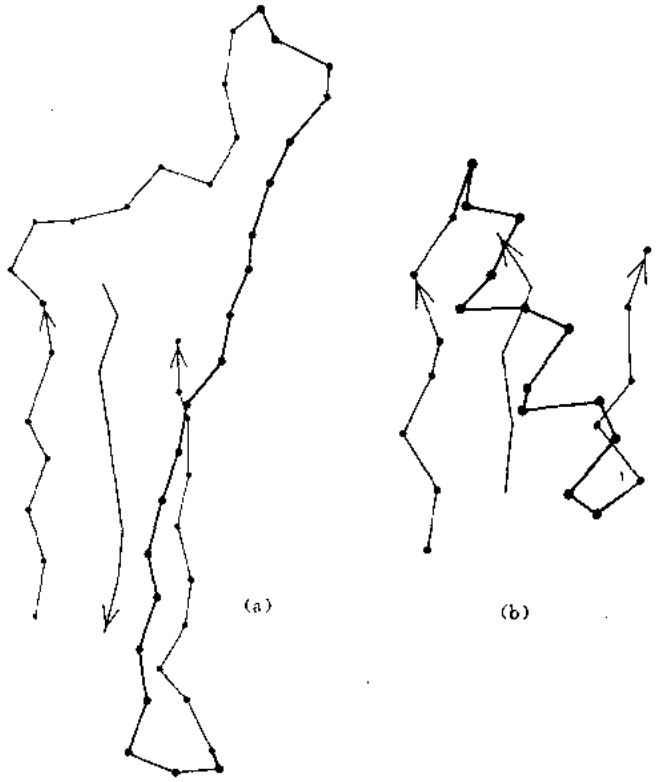


图 7.24

(a)反平行β链间的右手连接(伴刀豆球蛋白 A 残基169—214);

(b)平行β链间的α螺旋右手交叉连接(羧肽酶 A 残基200—242)

是由两个接近于直角的转折使肽链折回180°,残基 n 和 $n+3$ 的 C_α 相距仅0.5nm。类型Ⅱ的4个残基都具有相同的构象角,即 $\phi_1=\phi_2=\phi_3=\phi_4=-60°$,$\psi_1=\psi_2=\psi_3=\psi_4=-30°$,这种结构实际上就是重复结构$3_{10}$螺旋。

在已知三维晶体结构的蛋白质中,发现确实存在各种类型的β转角,见图7.30。除了典型的类型Ⅰ和Ⅱ各有一个氢键外,有些非典型的转角有两个氢键或者无氢键,后者可能是靠邻近侧链间的作用力或远程作用力来维系的。

β转角是球状蛋白质的一种重要二级结构。首先,它使肽链的方向发生变化。通常认为β转角在蛋白质折叠成天然构象的过程中,对确定肽链的走向起决定作用。第二,β转角在序列上相邻的各种二级结构之间起连接作用,是α螺旋与α螺旋之间以及α螺旋与β折叠之间十分好的接头,特别是它常作为相邻的反平行β链之间的发夹连接。此外,β转角还常常存在于α螺旋的末端。因此,β转角在球状蛋白质分子中占有相当大的比例,例如在由85个残基组成的高电位铁蛋白中就有17个转角,而且它们都位于分子的表面,见图7.31。

亲水性残基具有最强烈的形成β转角的倾向,这是由于β转角通常位于分子表面的原因;此外,Gly 和 Pro 在β转角中也是常见的,已发现在β转角的中央两个残基中,大约有2/3是 Pro-Asn 或 Pro-Gly。

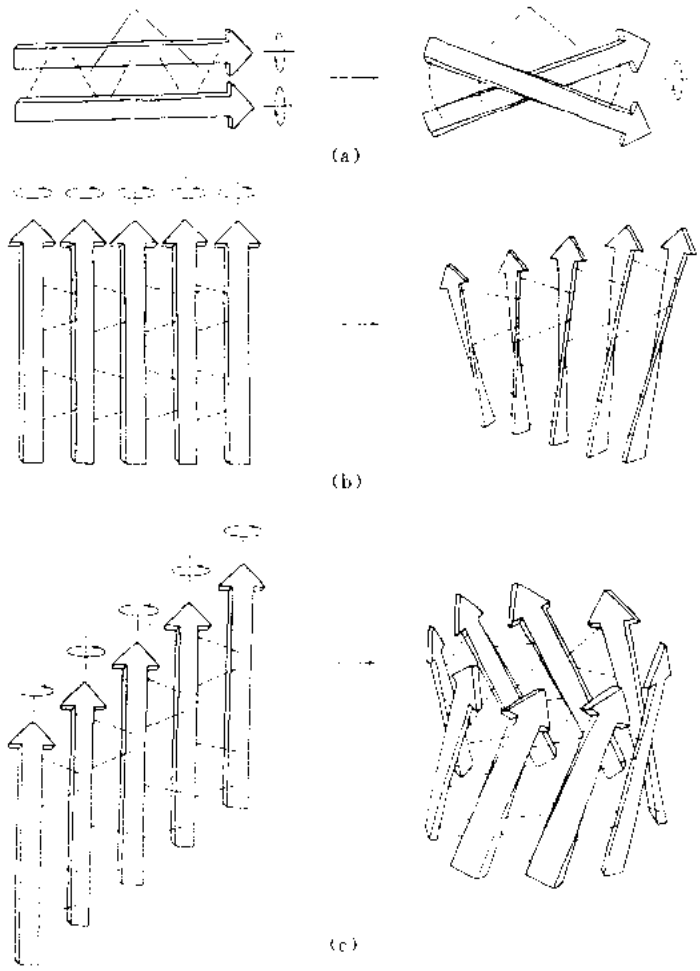


图7.25　平行β折叠片层的示意图

(a)右手扭曲的片层;(b)右手扭曲的马鞍型片层;(c)右手扭曲的β桶

7.4.2.4　β突起(β-bulge)

β突起是β折叠结构中的两个连续的氢键之间的一个突起的区域,它包括了在一条β链上的两个残基和在相邻的另一条链上的一个残基。图7.32为各种β突起的示意图。典型的β突起最常出现在反平行的β结构的窄空间的一对氢键之间,而且残基1,2和 x 的侧链都在β片层的同一侧。残基1的构象角平均为 $\phi_2=-100°$,$\psi_2=-45°$,接近于α螺旋的构象;残基2和 x 的构象角平均为 $\phi_2=-140°$,$\psi_2=160°$,$\phi_x=-100°$,$\psi_x=130°$,接近于通常的β折叠构象。图7.33为β突起的实例。

β突起不如β转角那么普遍,但是,β突起像β转角一样,改变了肽链的方向,不过,β突起是以缓慢得多的方式实现的。β突起经常出现在蛋白质的活性部位,因此,人们猜测β突起的另一个可能的功能是,在不破坏β折叠片层整体结构的情况下,对突变体中插入或删除一个残基起调节作用。

7.4.2.5　卷曲或无规卷曲(coil 或 randon coil)、紧密"环"(compact loops)和"连接"

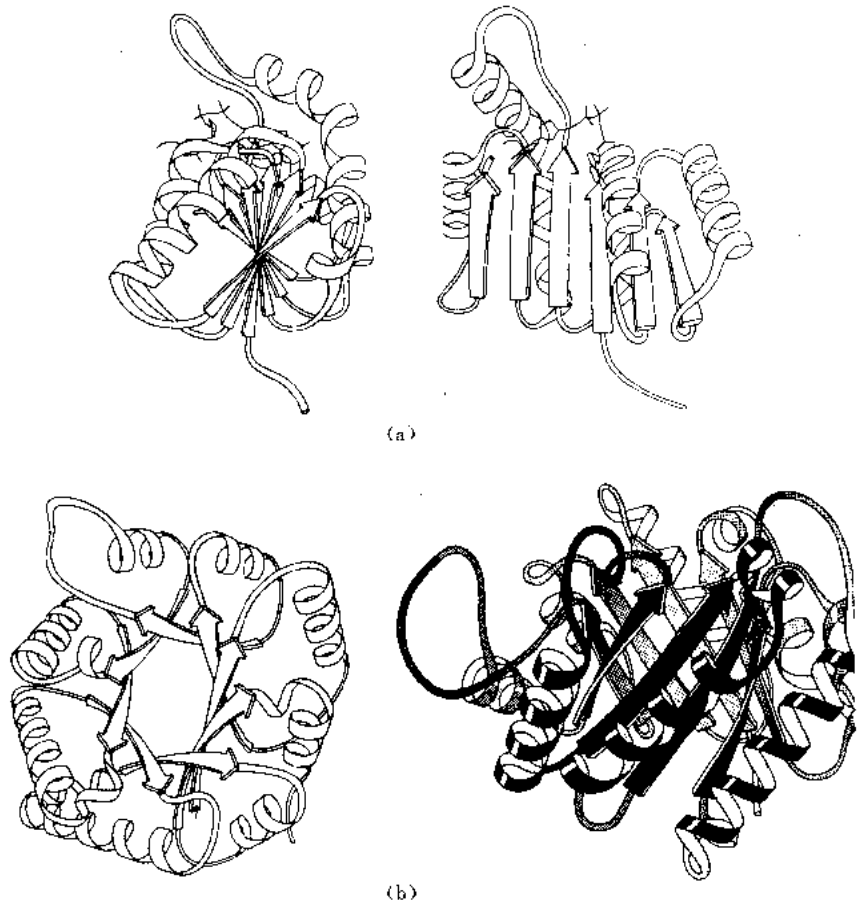


图7.26　平行β折叠片层的右手扭曲实例
左为顶面观，右为侧面观。(a)马鞍形(乳酸脱氢酶结构域1)；
(b)β桶(磷酸甘油醛异构酶)

(connections)

在蛋白质中，右手α螺旋和伸展的β链是仅有的两种具有重复的ϕ和ψ角度的构象。但是，蛋白质结构中的大部分是由十分有序的但不重复的构象所组成的，通常称为卷曲或无规卷曲。从名称看，似乎它具有无序的、可运动的、不折叠的涵意，但实际上它是有序的、紧密的、而且是与重复的二级结构片段一样稳定的。由于它是不重复的构象，因而不仅具有主链氢键，而且常有侧链与主链之间的氢键。无规卷曲包括相当不同的两大类：即各种紧密"环"以及二级结构的各个片段之间的各种"连接"。β转角可以看成是最短的环和连接，但在较大的环或连接中也可能包含β转角。

紧密环是指一个连续的肽链片段在三维空间中形成的一种环状构象。此片段的两个末端之间的距离很小，因而一个理想化了的环的主链折叠像一个希腊字母Ω，因而又称为Ω环(Ω loops)[46]。与主要以主链结构来定义和使之稳定的二级结构如α螺旋、β折叠、β转角等不同，环是以侧链来定义和使之稳定的，即是以侧链间的相互作用而形成紧密的环，而且大多数氢键是由侧链形成的；它们的主链扭角是非重复性的，而且主链NH与C=O间的氢键即使有也是

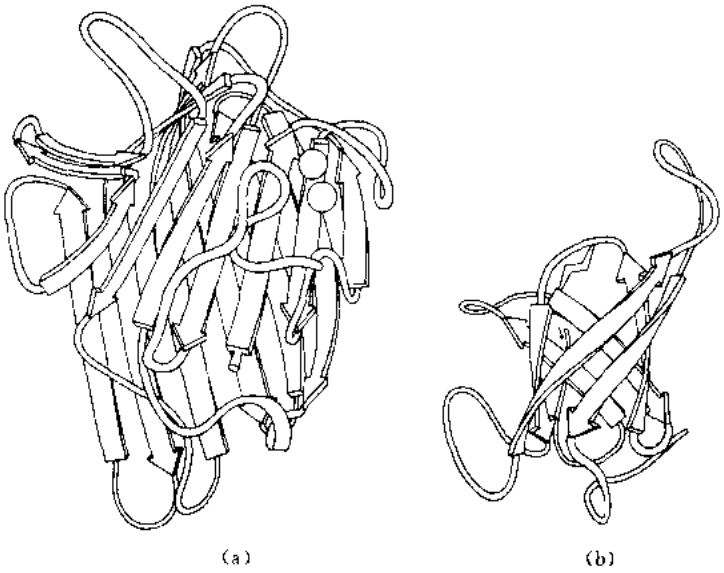


图7.27　反平行β折叠片层的右手扭曲实例
(a)部分扭曲的(伴刀豆球蛋白 A);(b)β桶(胰蛋白酶结构域1)

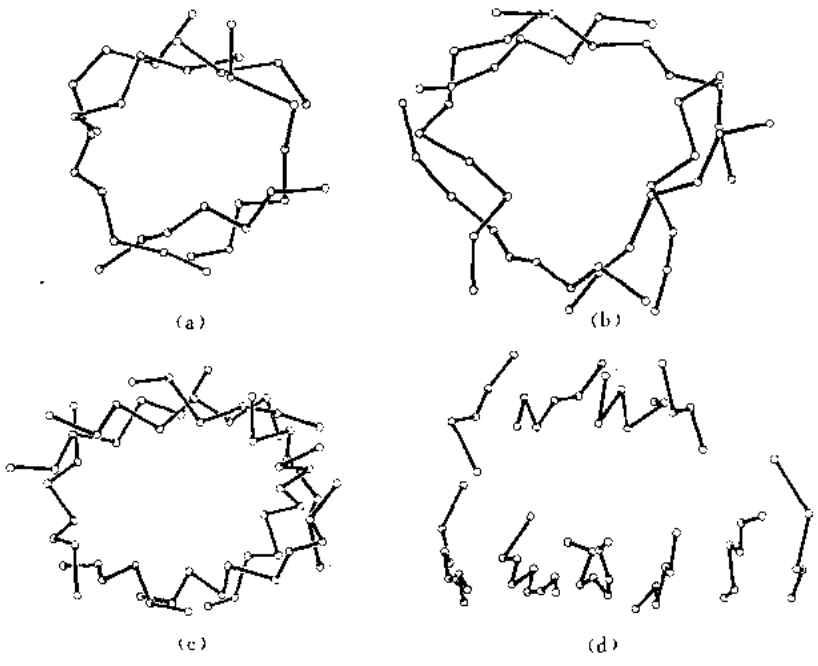


图7.28　β桶的顶面观实例

很少的。简单的Ω环是以下述三个标准来定义的:(1)片段的长度必须是6—16个残基。若短于6个则为β转角。从表面上看,一个β转角只不过是一个小的环,但是由于转角长度为3—5个残基,因而主链基团紧密地堆积在一起,致使侧链朝外伸出。而环的片段较大,无此立体化学限

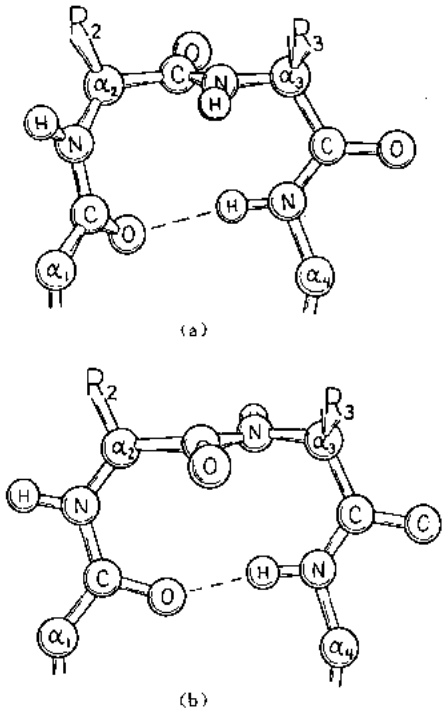


图7.29 β转角的两种主要类型示意图
(a)类型Ⅰ(普通转角);(b)类型Ⅱ(甘氨酸转角)

制,因此侧链堆积在环内部的核心中。环的长度的高限定义为16是根据大多数实际的环所具有的最大残基数。(2)是无规则的二级结构。由这一标准即可将邻近的β折叠的链和其他不确定的结构与其区分开来。(3)片段两末端之间的距离,即从片段的第一个 C_α 到最后一个 C_α 之间的距离,必须小于1nm,而且不超过在所考虑的片段中任何两个 C_α 之间的最大距离的2/3。以这一标准定义的片段像一个Ω形状的环。实际观察到的环的两末端间的距离为0.37—1.0nm。图7.34为典型的Ω环的实例。

Ω环几乎总是位于蛋白质分子的表面,而且在具有多个Ω环的蛋白质中,这些独立的环在空间中常常成簇地分布,见图7.35。

Ω环的氨基酸组成表明,在β转角中最常存在的残基也常在环中出现,即 Gly,Pro,Asp,Asn,Ser。此外,在Ω环中还常出现 Tyr。可以看到,除了 Tyr 以外,所有这些残基都具有短的侧链,而且除了 Pro 以外都是极性的、亲水性的。

由于环几乎总是位于蛋白质分子的表面,而且具有特殊的构象形式,因而承担了重要的生物学功能,特别是生物学的识别功能以及分子进化的突变部位。

在蛋白质中,Ω环是普遍存在的。在小蛋白质分子中可能无环;在较大的蛋白质中,平均每个分子具有大约4个环,环中残基约占分子残基总数的21%。虽然环在经典的二级结构类型中是属于无规卷曲,但事实上它们是高度紧密的亚结构,而且还可能是独立的折叠单位。

在蛋白质中,规则的二级结构片段之间的"连接",除了前面已经讨论过的β转角和β突起常常成为反平行β链之间的发夹"连接"以外[见图7.36],还有一些相当伸展的"条带"(straps)

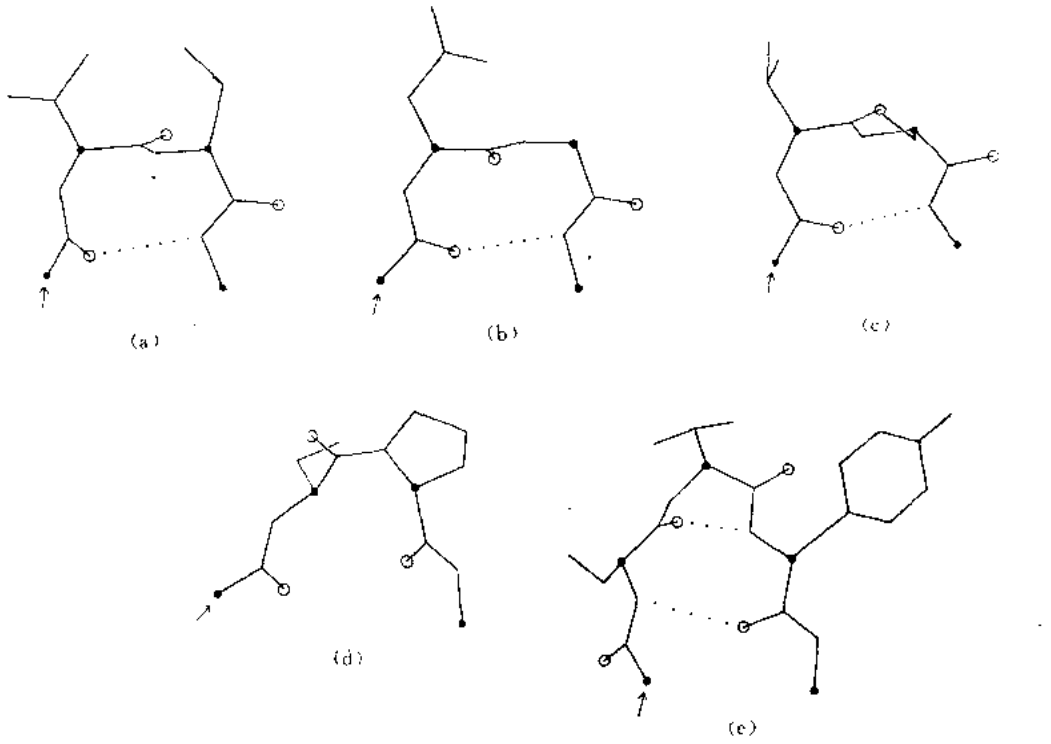


图7.30 β转角的实例

(a)类型Ⅰ;(b)类型Ⅱ;(c)类型Ⅲ;(d)无氢键的;(e)有两个氢键的

图7.31 高电位铁蛋白的主链结构

粗线表示β转角的中央肽

也成为各种二级结构片段之间的"连接"。按照所连接的二级结构类型和相对位置,可将"条带"分成不同的类型:(1)短的反平行的和短的垂直的 α-α 连接。短的 α-α 连接当从末端观察时常具有"S"形状,见图7.37。(2)β-α 和 α-β 连接。它们具有很不相同的长度和构象,其中一些短的连接具有相近的长度和构象以及残基种类。(3)反平行 β-β 发夹连接。除了 β 转角和突起以外,另一种类型的发夹连接是所谓的"希腊花边连接"(Greek key connection),又称为 β 弓(β-arch)。它连接一个 β 片层的同一端的不相邻的 β 链,并常常在 β 桶的一端形成弓形的连接,见图7.38。最常见的"希腊花边连接"是具有4 -5个残基的条带型连接。

随着高分辨率蛋白质晶体结构资料的迅速增加,必然还将在无规卷曲中发现新的构象类型。

7.4.2.6 无序结构(disordered structure)

无序结构是指没有确定的空间结构的区域,在 X 射线衍射分析法所得到的电子密度图上是完全看不到的或是比蛋白质的其他部分的电子密度较为弥散的区域。无序结构是相当普遍地存在的,有的是由于分子的运动引起的;有的是由于在蛋白质晶体中各个分子的这个区域具

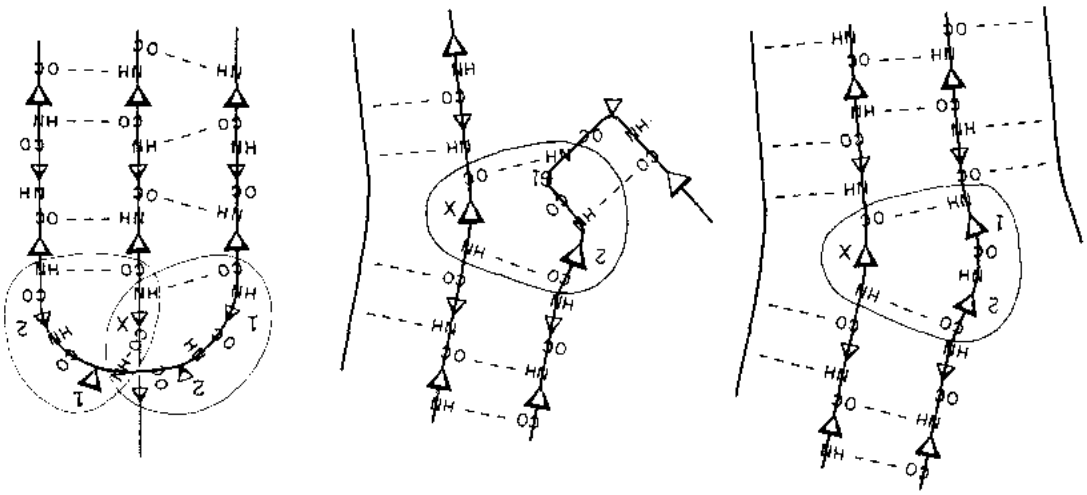


图7.32　各种类型β突起的示意图

大小三角形分别代表在片层上面和下面的侧链

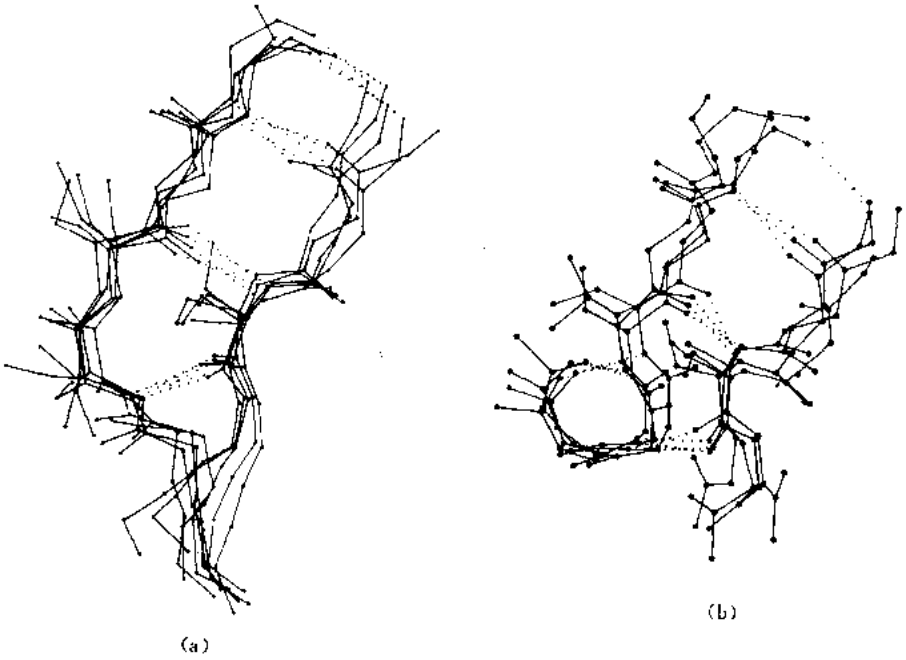
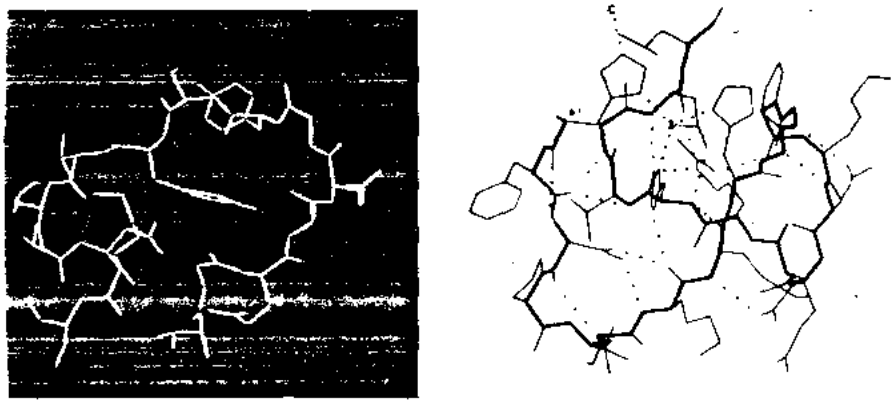


图7.33　β突起的实例

(a)5个典型的β突起的叠合比较;(b)4个与β转角类型Ⅱ(右下方)相连接的非典型的β突起的叠合比较

有不同的构象所致,此时可表示为不同的特殊构象的混合物。许多暴露于溶剂的带电荷的长的侧链,特别是 Lys,Arg,Ile 等常常是无序的,见图7.39。另外,多肽链的 N 末端和 C 末端的几个残基,假若不包含有大的疏水残基的话,也常常是无序的。



(a)　　(b)

图7.34　Ω环的实例

侧链堆积在内部，形成氢键，主链围绕在外部。(a)细胞色素c(残基40—54)的一个典型的环；(b)铜锌超氧化物歧化酶的一个大的紧密环

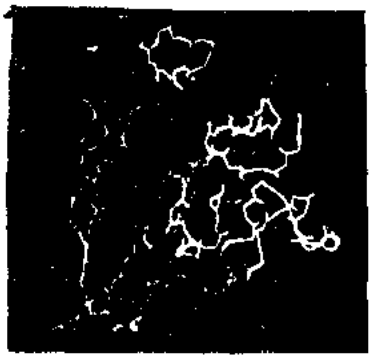


图7.35　铜锌超氧化物歧化酶中的一些紧密环成簇地分布(粗白线)

无序结构在一些情况下是没有重要生物学意义的，例如有些实验表明，切掉肽链末端的无序部分的残基，并不引起蛋白质活性的丧失。但在有一些情况下，无序区域具有十分确定的生物功能，如在某些配体结合位置处，当配体不存在时，肽段是完全或部分无序的；而当结合了配位体后，就成为十分有序的了。

7.4.2.7　结构水

与蛋白质分子结合的水分子实际上是蛋白质结构的一部分，与大量的无序的水分子不同，在高分辨率电子密度图中可以精确地确定它们的位置，见图7.40。结构水分子对维持蛋白质三维结构的稳定性起着一定的作用，尤其是对于暴露在外部的侧链的构象以及二级结构的末端的稳定性起着决定性的作用。有序性最好的水分子是结合在带电荷的侧链周围的水分子和存在于蛋白质分子表面的缝隙中的水分子，它们常常形成氢键网的一部分，见图7.41。另一些水分子埋藏在蛋白质内部，填充侧链间的空隙，并常常与蛋白质分子形成氢键，使结构稳定，见图7.42。此外，结构水分子有时甚至占据了活性中心的部分空间，与蛋白质的生物功能有着密切

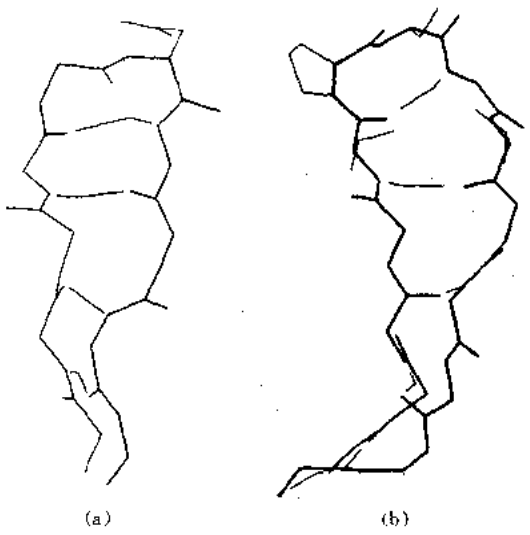


图7.36　相邻β链之间的发夹连接实例
(a)β转角作为连接(铜锌超氧化物歧化酶残基20—29)；
(b)β转角加β突起作为连接(Actinoxanthin 残基68—81)

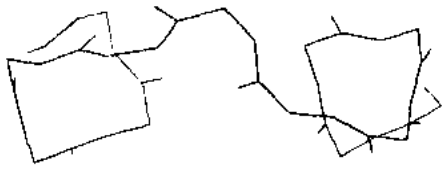


图7.37　短的α-α连接的实例
蚯蚓血红蛋白(残基33—45)形成"S"形连接

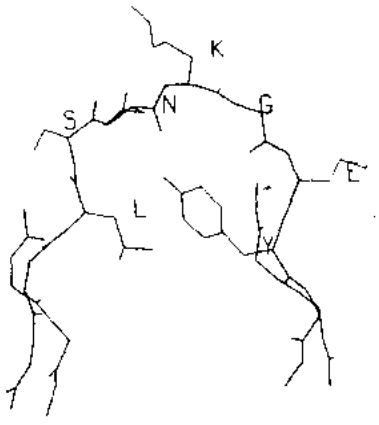


图7.38　由5个残基形成的一个"希腊花边连接"的实例

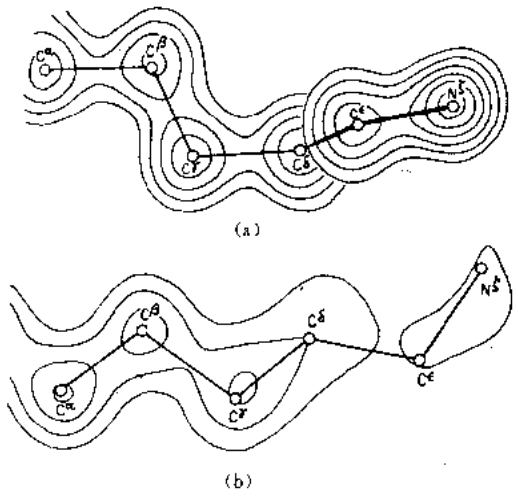


图7.39　无序结构的实例——红素氧还蛋白0.12nm 分辨率电子密度图

(a)有序的 Lys46；(b)无序的 Lys3

的关系。例如胰蛋白酶的底物的特异性水解位点 Arg 或 Lys 带正电荷的侧链，是通过水分子 H_2O414和 H_2O416与胰蛋白酶活性中心的底物结合口袋中带负电荷的 Asp 形成氢键而结合的，见图7.43。

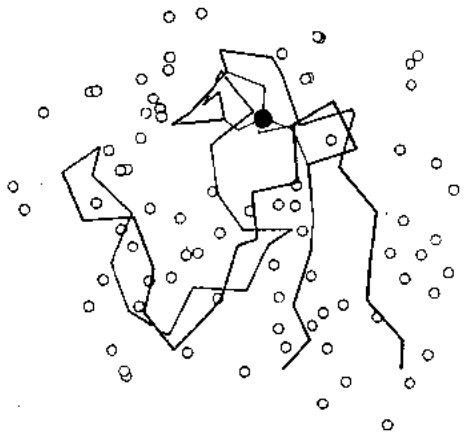


图7.40　蛋白质的结构水

红素氧还蛋白0.12nm 分辨率主链结构，o 代表水分子

7.4.2.8　二硫键在蛋白质空间结构中的重要作用

二硫键—S—S—属于共价键，因而应属于一级结构讨论的范围。而且，蛋白质通常能在没有二硫键或在二硫键形成以前正确地折叠成确定的空间结构。但是，二硫键的形成产生了局部的特异性，并增加了蛋白质的稳定性。因此，实际上二硫键是更多地以二级结构单元的作用而影响蛋白质的空间结构的。

二硫键的各个构象角的定义见图7.44。所有资料表明，二硫键的 C_α—C'_α 距离为0.4—

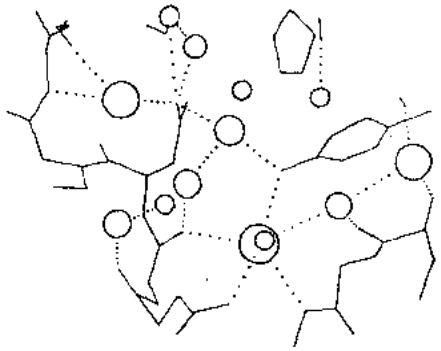


图7.41　蛋白质分子表面的水分子氢键网

0.12nm分辨率红素氧还蛋白的结构，图中圆圈的大小代表水分子占有率的高低，即有序性的大小

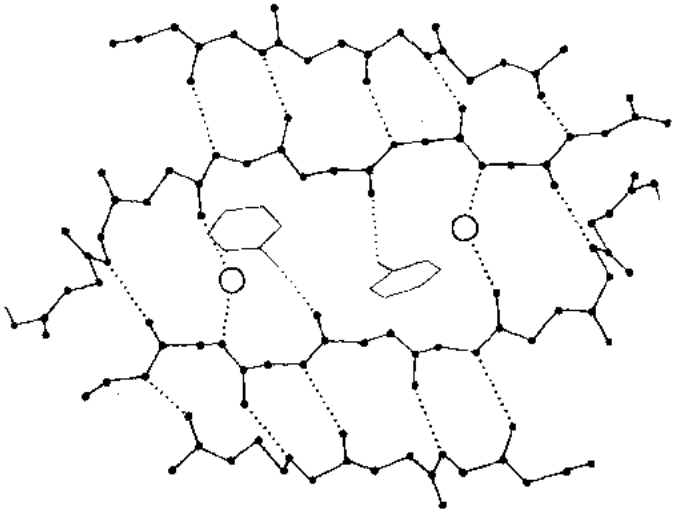


图7.42　结合水分子在β折叠片层主链基团之间形成氢键

0.9nm，其中大多数在0.44—0.68nm之间。无论在小分子中或在蛋白质中，都存在有右手和左手构象型式的二硫键，见图7.45中蛋白质的二硫键实例。大多数右手型和左手型的C_α—C_α平均距离和各个构象角的近似值见表7.3。

表7.3　二硫键的构象

	C_α—C_α平均距离(nm)	χ_1(°)	χ_2(°)	χ_3(°)	χ_2'(°)	χ_1'(°)
右手型	0.52	−60	+120	+90	−50	−60
左手型	0.61	−60	−90	−90	−90	−60

二硫键在蛋白质分子中的分布和位置是无规律的，但是两个半胱氨酸之间至少有两个残

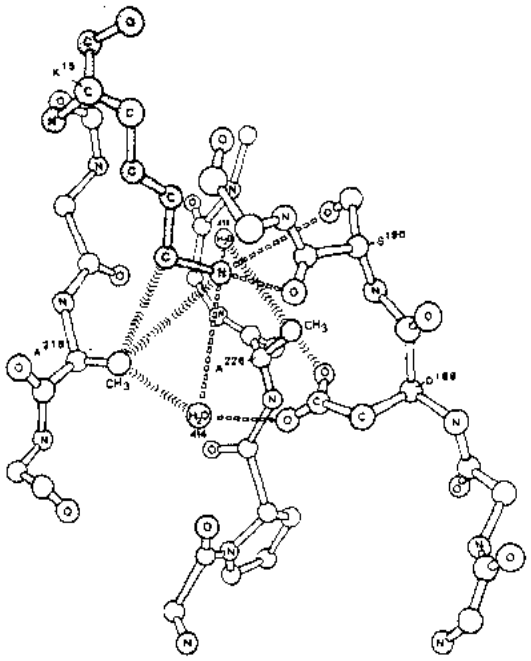


图7.43　胰蛋白酶活性中心的底物结合口袋中的结构水分子
胰脏胰蛋白酶抑制剂Lys15通过水分子414和416与酶形成氢键

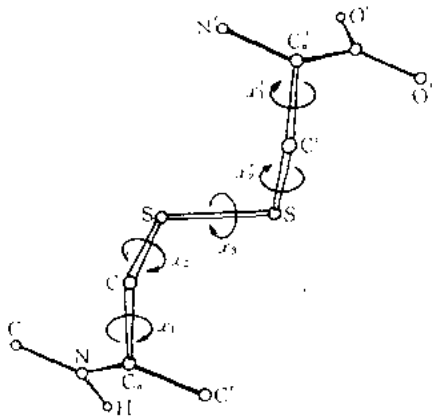


图7.44　二硫键的构象角定义

基相间隔才能形成二硫键；二硫键的两端大多为伸展的构象，其中有的实际上是β折叠片层的一部分；二硫键一般不可能连接同一个β片层中的相邻肽段；二硫键的一端连至一个α螺旋，另一端连至一条伸展的链或β转角或末端不规则的α螺旋，也是十分普遍的；但很少为两端都连至α螺旋；二硫键末端的另一种常见的主链构象是一个β转角，有时二硫键是转角的继续。

7.4.3　球状蛋白质的三级结构

蛋白质的折叠过程是将多肽链排列成一个正确的三级结构。也就是说，蛋白质的三级结构

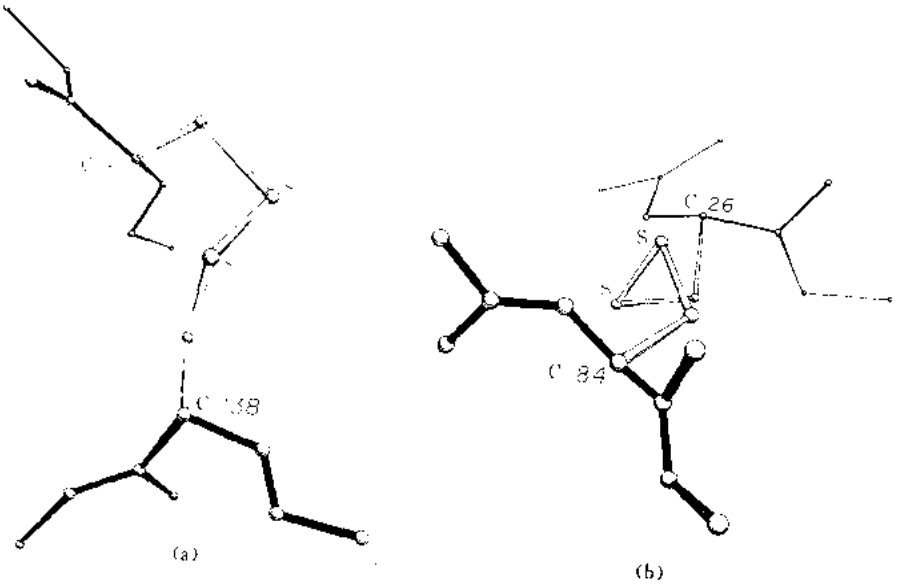


图7.45　二硫键的实例

(a)右手型(羧肽酶 A);(b)左手型(核糖核酸酶 S)

是多肽链的最后折叠状态。虽然从蛋白质的晶体结构中看到了千变万化的三级结构,但是却是十分有规律的。首先,各种二级结构单元以一定的方式结合形成较大的折叠单位以及更大的结构域;而且,所有这些已知晶体结构的蛋白质分子的结构域或整体三级结构都可归属于几种折叠类型之中。

7.4.3.1　标准折叠单位(standard folding units)

球状蛋白质的三级结构中存在着某些标准折叠单位。若将各种二级结构单元比喻成化学中的"元素",那么标准折叠单位则是化学中的"根"或"基团",因而又被称为超二级结构(super secondary structure)。也就是说,两种主要的二级结构单元即 α 螺旋和 β 折叠常常相互连接排列成多种确定的折叠类型。常见的三种标准折叠单位表示为(αα)、(ββ)以及(βαβ),见图7.46。(αα)单位是由两个相邻的 α 螺旋形成的,两者之间由一条肽链相连;(ββ)单位是由 β 折叠链形成的,连接处可为发夹结构;(βαβ)单位也是由 β 折叠链形成的,不过是由一个不规则的环链或一个 α 螺旋将 β 折叠的两条平行链的末端连接起来。图7.46中以示意图和拓扑图两种方式表示这些折叠单位以及蛋白质三级结构的特点。

拓扑学作为几何学中的一个分支,它仍然是研究图形或形状的科学,它研究的是图形在连续变形情况下的不变的整体性质。用拓扑图描述蛋白质的三维结构,虽然不十分清晰,但是其优点在于保留了较多的三维结构特点,是一个整体图形。它用一个二维的平面图代表了三维的立体结构,而若用三维结构图,则需用许多个不同的图才能描绘一个立体结构。用一个简单的不含糊的拓扑图可以描述各二级结构单元之间的连接情况和三级结构中肽链的走向等结构特点,虽然在文献中可以看到以不同方式表示的这种拓扑图[47,48,49]。每个拓扑图的形式和名称是按照每个连接跳过多少个 β 股或 α 螺旋以及连接的方向而定的,而与连接环的结构种类和构象无关。例如用"+1"表示连接相邻两条 β 链的发夹连接,如图7.23(a);若为交叉连接,则加一个字母"*X*"。例如图7.23(b)中连接相邻两条 β 链的交叉连接,表示为"+1*X*";跳过一条 β 链的

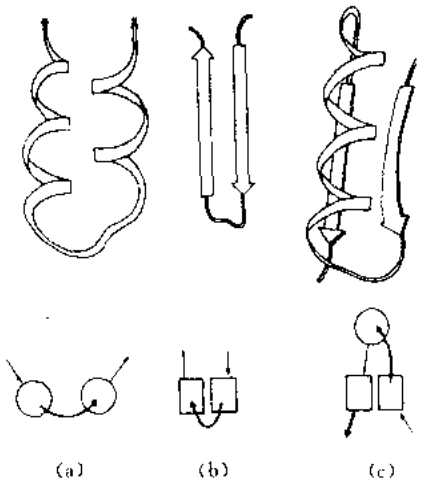


图7.46　标准折叠单位

(a)(αα)；(b)(ββ)；(c)(βαβ)。上图为带形示意图，螺旋代表α螺旋，箭头代表β折叠链；下图为拓扑图，□代表β链，○代表α螺旋

发夹连接或交叉连接则分别表示为"+2"或"+2X"等等。对于一个具有 n 条β链的β片层来说，则从N末端开始，逐个列出它的所有的连接，直至第 $n-1$ 个连接。例如图7.47(a)的黄素氧还蛋白可描述为"+1X，−2X，−1X，+1X"或"−1X，+2X，+1X，+1X"，其中正或负符号是没有意义的，因为片层可以翻转过来看，还可以用一个简化了的拓扑图来描述它的整体结构，见图7.47(b)。

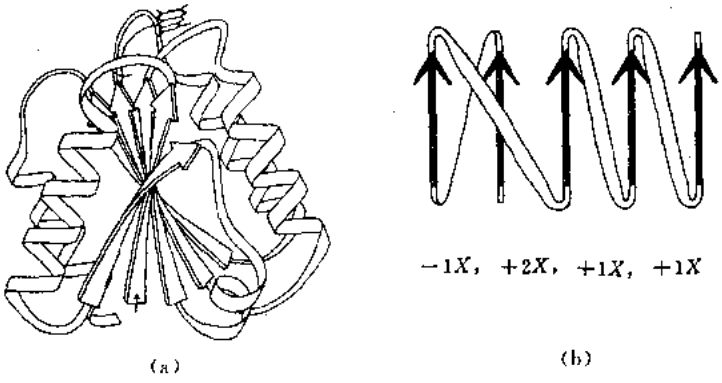


图7.47　黄素氧还蛋白的结构、简化拓扑图和它的名称

在上述折叠单位中，螺旋-螺旋接触时，一个螺旋表面的隆起总是对着另一个螺旋的槽沟，并且螺旋轴相互形成一定的角度；螺旋-β折叠链相互接触时，常常是两者的轴接近于平行；β折叠链-β折叠链相互接触时，决定于它们的扭曲程度。

在球状蛋白质中所观察到的所有三级结构，都不是杂乱无规的，而是可以用上述这些折叠单位以及它们的组合型来描述的。

7.4.3.2 结构域(domain)

在较大的球状蛋白质中，一条多肽链经常折叠成几个各自局部结构紧密的、半独立的球形结构区域，称为结构域。一般来说，一个结构域可以包含在序列中为连续的40—400个以上的残基，最常见的为100—200个残基。图7.48中的一些实例表明了在一个蛋白质分子中，可以包含2个、3个或多个结构域；各个结构域可能是相似的，也可能是完全不同的结构类型；结构域之间的连接，有的是十分紧密的，有的是由一条肽链形成的柔软的连接；结构域之间常常形成裂隙，蛋白质的活性部位和酶的底物结合部位常常在裂隙中。虽然关于结构域的定义[43,50,51]有很不相同的描述方式，特别是在关于结构域究竟能小到什么程度这个问题上还有一些争执，但这些分歧并不重要。

以笔者[35]参加分析的大肠杆菌麦芽糖结合蛋白为例[见图7.48(a)]。它是由370个氨基酸残基组成的，分子量为40622道尔顿。其分子结构呈线度为6.5nm×4.0nm×3.0nm 的椭圆形，它是由两个球状的结构域所组成的，两者由一个1.8nm 深、底部线度为0.9nm×1.8nm 的缝隙或沟分开，将这两个结构域分别称为 N 结构域(含有 N 末端)和 C 结构域(含有 C 末端)。N 结构域(左)由肽片段1—109和272—319组成；C 结构域(右)由肽片段114—267和334—370组成，它可以进一步分成两个部分，较大的部分称为 C_1结构域(右上方)，较小的称为 C_2结构域(右下方)。N 结构域与 C 结构域具有十分相似的二级结构元素的排列(即超二级结构)，都具有一个5股的中央β片层核心，在片层的两侧都各具有2个或3个α螺旋。两个结构域之间仅由3段肽链相连接，形成缝隙。麦芽糖埋在分子中部的这两个结构域之间的缝隙中，以氢键和范德华力相互作用而与该蛋白质相结合，并被运输至菌体内。

图7.48(b)为丙酮酸激酶。可以看到，它具有3个结构完全不同的结构域。

结构域的另一个最重要的生物学功能是运动。结构域有时能作为独立的刚体而运动，结构域之间柔性的连接对于底物的结合、变构的控制以及分子整体的组装等都具有决定性意义，例如图7.48(c)中的己糖激酶结构。当它未结合葡萄糖分子时，酶分子的底物结合部位的局部构象较为松弛，而当它结合了葡萄糖以后，则变得较为紧密。结构域的再一个重要功能可能是作为分子的建造单元以利于有效的组装。可以假定，在蛋白质生物合成过程中，各个结构域独立地折叠起来，然后迅速组装成分子整体，这样，加速了蛋白质大分子的折叠过程以适应生物学的要求。

关于结构域的概念，除了上述有关结构紧密性的纯几何学概念以外，还有另外两个相对独立的概念。一是在蛋白质被水解或酶解成各结构域肽链片段时，各结构域仍能折叠成稳定的结构，甚至还保留了生物活性；另一个是遗传学的概念，一些蛋白质的结构域的分割与 DNA 中的外显子十分精确地符合。因此，结构域不仅仅是折叠的单位，而且还可能是结构和功能的单位以及遗传的单位。最有趣的是，在一些具有不同功能的蛋白质中具有相同的结构域。例如所谓“核苷酸结合结构域”(nucleotide binding domain)，见图7.52。这说明这些蛋白质，至少它们的一个结构域，是由一个共同的祖先进化而来的。

7.4.3.3 三级结构的折叠类型

虽然蛋白质以及它们的结构域的三级结构多种多样，但是仍然存在着一些规律，可以将它们归纳为几种折叠类型。由于同一个蛋白质中的各个结构域常常是完全不同的，因此若以结构域为单位来对蛋白质的三级结构的折叠型式进行分类，问题就变得简单得多。到目前为止，根据每个结构域中二级结构的类型、数量和组合，主链骨架的各层相对于疏水核心的数目和形

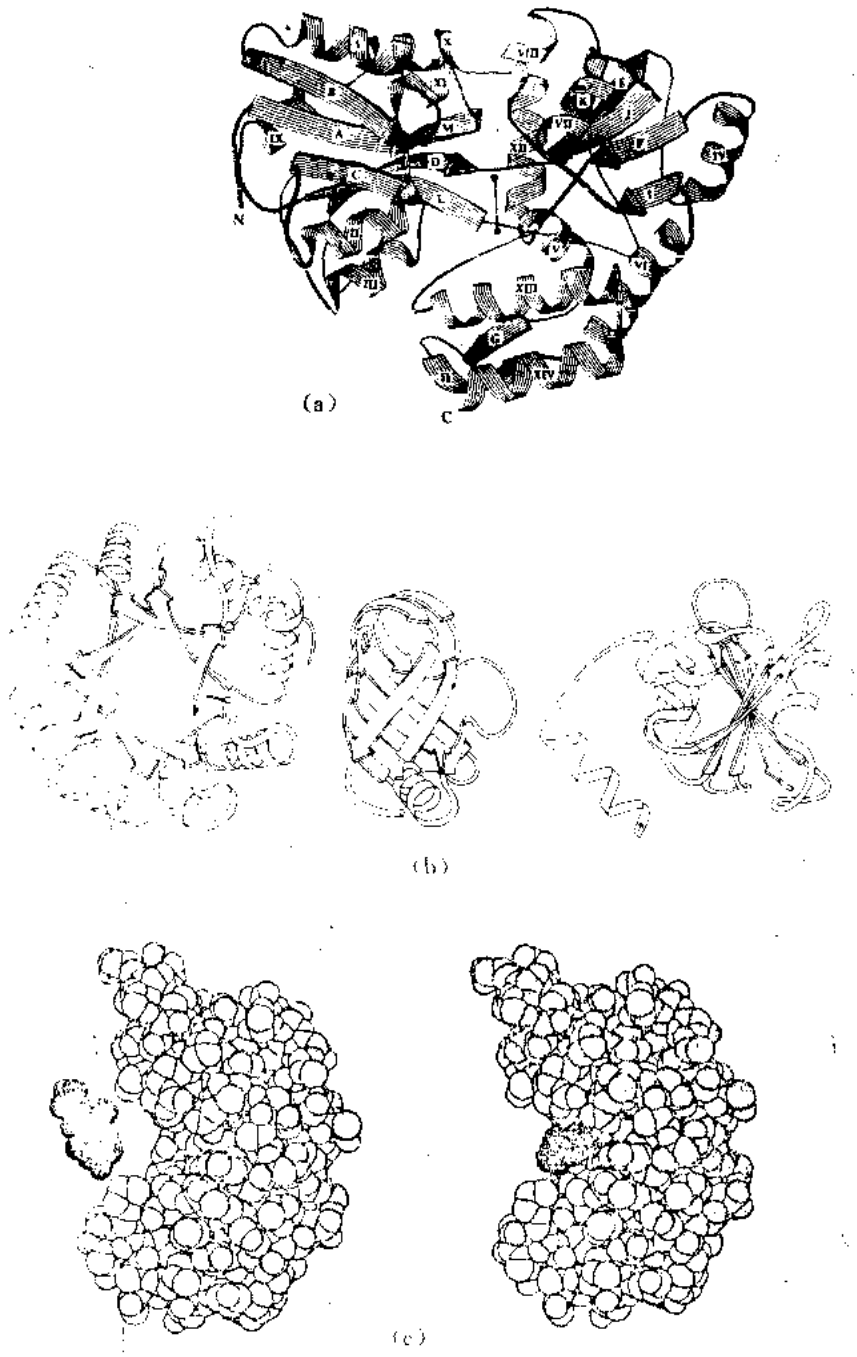


图7.48　各种结构域的实例

(a)麦芽糖结合蛋白，▮代表麦芽糖分子；(b)丙酮酸激酶的三个结构域；(c)己糖激酶，深色为葡萄糖分子

状，以及在一层中肽链之间连接的拓扑图等，可将结构域的三级结构分成四种主要类型，即反平行 α，平行 α/β，反平行 β 以及不规则的小结构四类。大约3/4的已知蛋白质结构可以用这四种类型加以描述。

第一类为全 α。这类三级结构是由 α 螺旋组成的，它们的 α 螺旋以平行的层或组排列起来，而且实际上它们总是为反平行的或接近于垂直的。全 α 类型中最简单的亚类是所谓升降螺旋捆(up-and-down helix bundle)，即相邻的 α 螺旋反向排列，头尾相连，形成近似于一个筒形的

螺旋捆或螺旋簇。最普遍的是4-螺旋捆，4个α螺旋形成稳定的两层的结构，如图7.49所示。至今只有几个3-螺旋捆的例子和一个2-螺旋对的例子以及极少数几个较大的螺旋捆的例子。在膜蛋白中，由于每个α螺旋都跨越膜，因而具有比球状蛋白质中更为精确地反平行的螺旋捆。

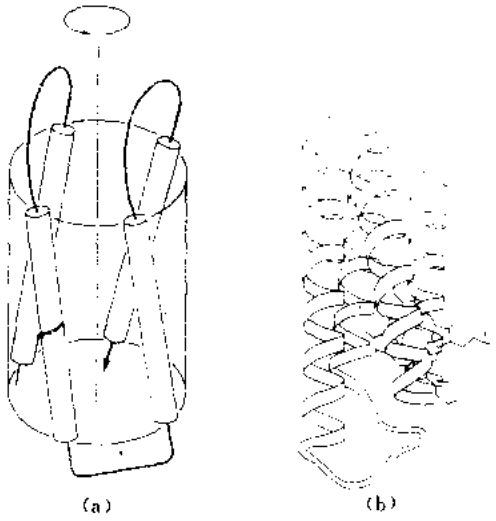


图7.49　反平行α螺旋捆的结构示意图(a)和实例：细胞色素 b_{562}(b)

上述这种由最靠近的相邻的二级结构单元相互折叠在一起的现象是蛋白质折叠的最常见的规律之一。这可能是由于从统计学上看，肽链的两个相邻片段比分离的片段折叠到一起更为迅速。全α类型的另一个亚类是所谓的希腊花边螺旋捆(Greek key helix bundle)，这是由于它很像希腊花瓶上常见的回形镶边。它不如升降螺旋捆常见。这种结构是由连续的螺旋对接近于相互垂直地折叠起来，并且大多以末端处相邻序列的残基相连接，形似回形花边，见图7.50。在全

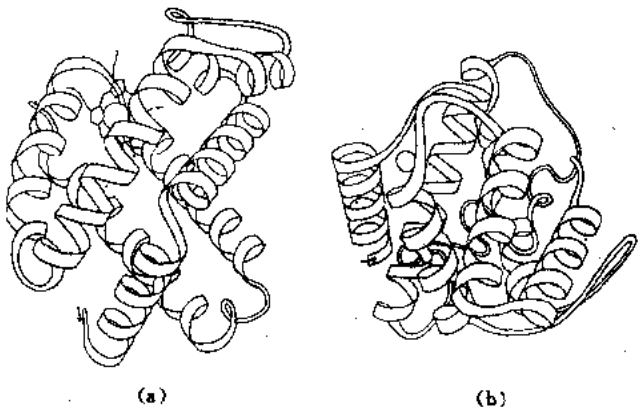


图7.50　希腊花边螺旋捆的实例
(a)血红蛋白β亚基；(b)嗜热菌蛋白酶结构域2

α蛋白质中，还有一些是各亚类的混杂型。全α蛋白质多为两层的结构，也发现有三层α螺旋的结构。

第二类为平行 α/β。这类结构是由 α 螺旋和 β 折叠链沿着肽链的方向交替存在，并折叠组成多层的结构。平行 β 折叠片层在内部，α 螺旋形成外部覆盖层。这种结构是由上节中所讨论过的标准折叠单位(βαβ)或交叉连接所建造起来的，而且绝大多数为右手交叉连接。平行 α/β 结构的第一个主要的亚类称为单绕平行 β 桶(singly wound parallel beta barrel)。在它的中央部位为一个由8条平行 β 链组成的 β 桶，并由8个 α 螺旋以右手交叉连接形式沿着同一方向逐个连接相邻的 β 链，形成与中央 β 桶同轴的外筒，见图7.51中的实例。平行 α/β 结构的第二个

图7.51 单绕平行 β 桶的实例
(a)丙酮酸激酶结构域1，(b)磷酸甘油醛异构酶

亚类称为双绕平行 β 片层(doubly wound parallel beta sheets)，或称为核苷酸结合结构域(nucleotide-binding domains)，见图7.52中的实例。这种结构域的中央部位是由4—9股完全平行的或接近于平行的 β 链所形成的扭曲的马鞍形的一层 β 折叠片层结构，α 螺旋在链间形成右手交叉连接，并在 β 折叠片层的两侧各形成一层 α 螺旋。β 折叠片层埋藏在内部，因而它是疏水性的，仅仅在末端和边缘的 β 股是亲水性的。这通常是平行 β 结构的特征。在这些结构域中，肽链通常从 β 折叠片层的中部起始，由 α 螺旋在最靠近的两个 β 股之间形成右手交叉连接并覆盖片层的一侧，然后改变方向返回到片层的中部并绕向覆盖片层的另一侧。核苷酸结合结构域是以 NAD 为辅基的各种蛋白质所共有的结构类型，在 β 折叠片层的 C 末端处有一个单核苷酸或二核苷酸结合位置，肽链在此处改变方向。具有这种结构的结构域或蛋白质，很可能是从一个共同的祖先歧化进化而来的。其中有许多是糖酵解途径中的酶。

第三类为反平行 β 结构。像平行 α/β 类一样，反平行 β 结构也能形成 β 桶和单片层，但是它们的拓扑图以及其他结构是不同于 α/β 蛋白质的。反平行 β 结构中最常见的亚类是所谓的“希腊花边 β 桶”(Greek key β-barrel)，见图7.53。这种结构的三维拓扑图摊平后的二维拓扑图很像希腊花瓶上常见的回形镶边，因而得名。所有这类结构的共同特点是至少有一条(可多至4条)链间连接是跳过相邻的 β 链因而是跨过桶的一端的。除个别例外，希腊花边 β 桶当从外部观看时，链是反时针方向盘绕的，而且在桶的相对着的两边之间为右手交叉连接。反平行 β 类中的另一个亚类是所谓的“开面夹心”结构(open-face sandwich)，见图7.54。它具有3—15股链的单层反平行 β 片层，虽也是扭曲的，但不封闭成桶形。在片层的一侧有一层 α 螺旋和一些环，通常在 β 片层与螺旋层之间为疏水侧链。在小蛋白质中，片层的另一侧是暴露于溶剂的，而在

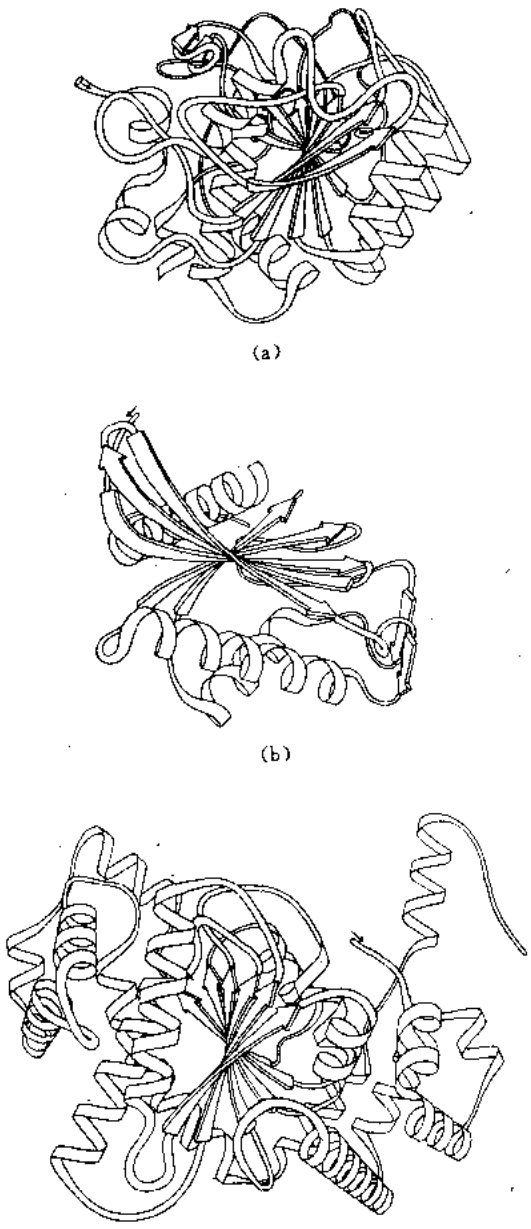


图7.52 双绕平行β片层结构(核苷酸结合结构域)的实例
(a)枯草杆菌蛋白酶;(b)己糖激酶结构域1;(c)磷酸化酶结构域2

较大的蛋白质中,它可能与其他结构或亚基接触。在反平行β类中还有一种所谓的升降β桶(up-and down β-barrel)亚类,见图7.55。在这种结构中,相邻链间具有发夹连接,类似于升降螺旋捆。不过,这种亚类结构是较少见的。

第四种类型是小的不规则的结构。许多很小的蛋白质是很难以归类到前述任何一种结构类型中的,它们的结构常常是不规则的,只具有相当少量的二级结构,而且不组织成较大的结构单位。它们常常是富含二硫键或半胱氨酸的,或者是富含金属或非肽基团的小蛋白质,见图7.56。但是,存在一个规律,即二硫键与金属或非肽基团是不相容的。因而将小的不规则结构分成富含二硫键和富含金属的两个亚类。这种限制很可能与肽链折叠时或折叠后,二硫键如何形成或非肽基团如何结合到蛋白质中去有关。在这类蛋白质中,除了在其他多数蛋白质中的二级结构单元之外,二硫键或非肽基团的几何学和堆积方式也是极其重要的因素。因此,若当二硫键被破坏或失去非肽基团时,这些蛋白质往往就不能保持天然三级结构。

7.4.3.4 球状蛋白质按三级结构分类

假若以三级结构中α螺旋和β折叠结构的存在与否以及它们的排列方式来对球状蛋白质进行分类,可以将它们分成4个主要类型。图7.57中列出了这4种类型的拓扑图实例。

第Ⅰ类,α:这类蛋白质由α螺旋以不规则构象的链连接而成。其中α螺旋不一定是平行的,它们通常形成两层α螺旋。

第Ⅱ类,β:这类蛋白质由β折叠结构组成,不含或只含有极少量α螺旋。它们通常形成两层β折叠片层。

第Ⅲ类,α+β:在这类蛋白质中,α螺旋和β折叠两种主要的二级结构单元都存在。它们在大多数情况下也是双层的。

第Ⅳ类,α/β:这是α螺旋和β折叠都存在的另一种类型。不过,在这类蛋白质中,α螺旋和β折叠的链是沿着肽链的方向交替存在的。这些蛋白质较大,结构较复杂,通常形成夹心的3层结构,β折叠在内部,α螺旋在两侧。甚至形成4层。看来,没有一个蛋白质为一层结构而是稳定的,因为至少要有两层才能埋藏疏水核。

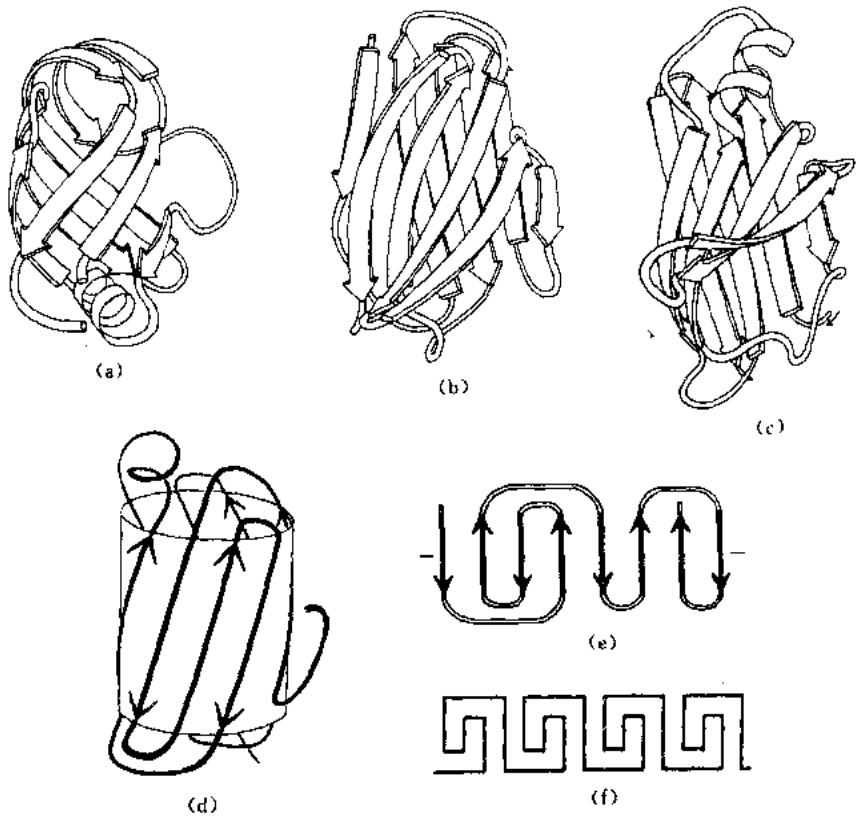


图7.53　希腊花边β桶亚类结构的实例

(a)丙酮酸激酶;(b)免疫球蛋白 V_L 结构域;(c)前清蛋白;(d)前清蛋白的三维拓扑图;

(e)将三维拓扑图摊开后的二维拓扑图;(f)希腊回形花边图案

7.4.4　球状蛋白质的四级结构

许多蛋白质分子作为一个活性单位和结构单位的整体来看,是由两条以上独立肽链组成的。这些肽链相互以非共价键(包括氢键、疏水作用和盐键等)连接成一个相当稳定的结构,实质上是一种缔合作用。这些独立的肽链称为亚基(subunit),多个亚基缔合成一个蛋白质分子。四级结构亚基的概念最初来自于病毒的蛋白质亚基。J. I. Harris 等人[52]用蛋白质末端基团测定方法发现烟草花叶病毒(TMV)具有许多相同的分子量约为17000道尔顿的蛋白质亚基。后来亚基的概念被推广应用于蛋白质分子中。两个亚基的缔合体称为二聚体,三个亚基的缔合体称为三聚体,同理有四聚体、五聚体、六聚体…等等。亚基的这种聚合的空间排列和结合的性质称为四级结构。一个蛋白质的亚基可以是相同的,也可以是不相同但密切相关的。亚基一般以希腊字母 α,β,γ,δ 等命名。例如血红蛋白分子包含 α 亚基和 β 亚基各2个,表示为 $\alpha_2\beta_2$(分子为四聚体),促甲状腺素为 αβ(分子为二聚体),神经生长因子为 αβγ(分子为三聚体)等等。具有四级结构的蛋白质相当多。蛋白质亚基的缔合、解离变化以及空间结构的变化与它们的生物学功能有着密切的关系。

用 X 射线衍射法和电子显微镜技术可以观察到蛋白质的四级结构。特别是 X 射线分析法已成功地精确测定了一些蛋白质的四级结构,为阐明四级结构与生物学功能的关系提供了论

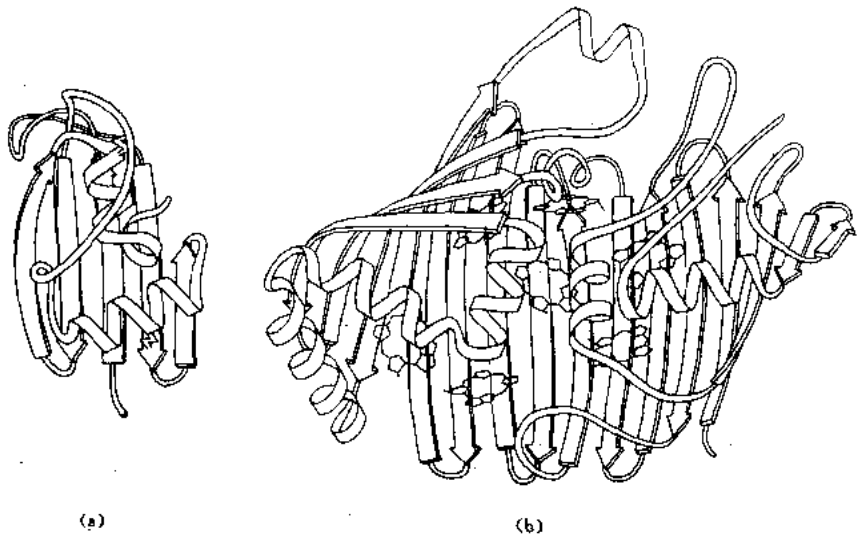


图7.54　开面夹心亚类结构的实例
(a)枯草链霉菌抑制剂;(b)细菌叶绿素蛋白

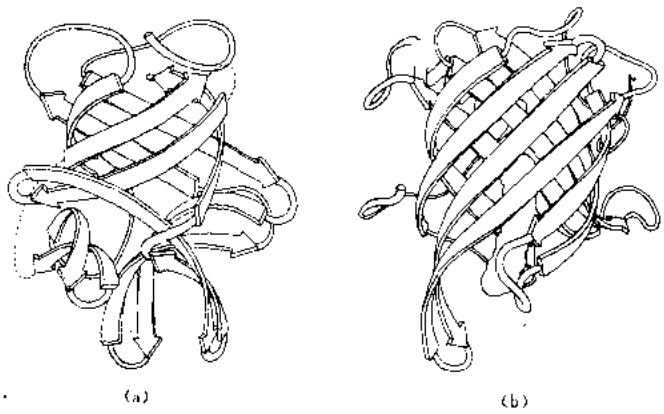


图7.55　升降β桶亚类结构的实例
(a)黄豆胰蛋白酶抑制剂;(b)视黄醇结合蛋白

据。

7.4.4.1　亚基的空间排列方式

蛋白质的四级结构是由若干个球形亚基高度有序地组织形成的一个具有生物学功能的聚合体。可以将它们分成两种不同的基本类型:

第一种类型是由各不相同的亚基组装的。它包括从不同类型亚基的二聚体分子直到多个不同类型亚基的复杂的组装如染色质核小体中的组蛋白八聚体(见第十章图10.5,10.6)以及核糖核蛋白体(它还含有核酸)等。在这种类型的四级结构中,各个不同的亚基与其相邻的每一个亚基之间的相互作用具有不同的特殊性质,而且这些相互作用只有在某一种聚集排列中才存在,因此这类整体结构复杂的四级结构具有高度不规则的几何学形状。

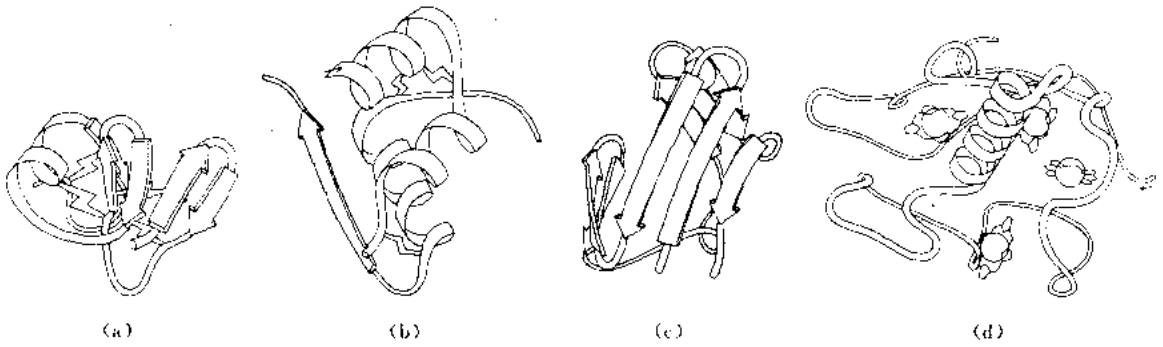


图7.56　小的不规则结构类型实例
富含二硫键的：小麦胚芽凝集素结构域2(a)和胰岛素(b)；富含金属的：红素氧还蛋白(c)和细胞色素 c_3(d)

第二种类型的四级结构是由一种或几种不同亚基的多考贝组装成的聚合物分子。由于亚基之间的特殊的结构相互作用重复出现，因而这些聚合物具有规则的几何学排列。这类四级结构是较为常见的。由于蛋白质是由手性的L-氨基酸所组成，因而是不对称的分子，所以四级结构最简单的花样可以是线性的聚合物。例如肌肉结构蛋白中的肌动蛋白和肌球蛋白等各自以线性缔合成细丝和粗丝，这两种肌丝在肌动蛋白和肌球蛋白的相互作用下产生收缩作用。又例如在血液凝固过程中，首先由血纤蛋白原变为血纤蛋白，然后血纤蛋白线性缔合成大分子而使血液凝固。在线性聚合分子中，相邻亚基之间特殊的结构相互作用重复出现，继续加入亚基可以延伸。比线性排列稍为常见的是相同亚基的螺旋排列。在这种四级结构中，各个亚基与相邻的各亚基之间具有不同的、局部的相互作用，但是这种相互作用对于每一个亚基来说是重复的。最明显的例子是棒状病毒和丝状病毒(见第九章图9.8和9.13)，完整的外壳是由相同的蛋白质亚基的多考贝组装成的，因而十分有效地利用了病毒核酸的信息量。与线性排列相似，螺旋排列的亚基沿着分子链上升具有重复的相互作用，因而它们可能是无限地延伸的。但是事实上，在螺旋形病毒中，外壳蛋白质的长度不是由蛋白质所决定的，而是由病毒核酸的固定长度决定的。除了线性的和螺旋的聚合体以外，许多蛋白质的亚基以各种对称性而规则地排列成的聚合体也属于这类四级结构，见图7.58中的实例。如本书作者之一参加测定的去六肽卵清蛋白分子[53]，是由同一种类的亚基以近似的二重旋转轴对称性排列的二聚体；又如本实验室参加测定的猪胰岛素二聚体和六聚体[42]；其中胰岛素二聚体中的两个单体以局部的非晶体学二重旋转轴相连系。三个二体再以晶体学三重旋转轴对称性排列成六聚体。又如以两对不相同但极为相似的亚基α和β以近似的222对称性排列成的血红蛋白四聚体分子，见图7.62。更复杂的规则排列的四级结构如球状病毒，它们的亚基(相同或不同)以近似的20面体对称性排列聚合形成病毒的蛋白质外壳，见第九章图9.14，9.15，9.17。在这类四级结构中，所有的亚基都是以几何学上相似的方式相互作用的。但是，令人惊奇的是，从X射线晶体学分析所得到的结果表明，四级结构中各亚基之间的相互作用往往是不对称的和不相等的，甚至在化学上相同的亚基之间也是如此。这种不相等的相互作用的最简单的类型是出现在一些二聚体的亚基接触处，即在靠近二重轴(实际上只是近似的)处的各个侧链，为了避免重叠，被强迫而处于不同的位置。有时这种非相等性只是局部的，可能与生物学功能无关；但是有时非相等性还延伸到亚基的其

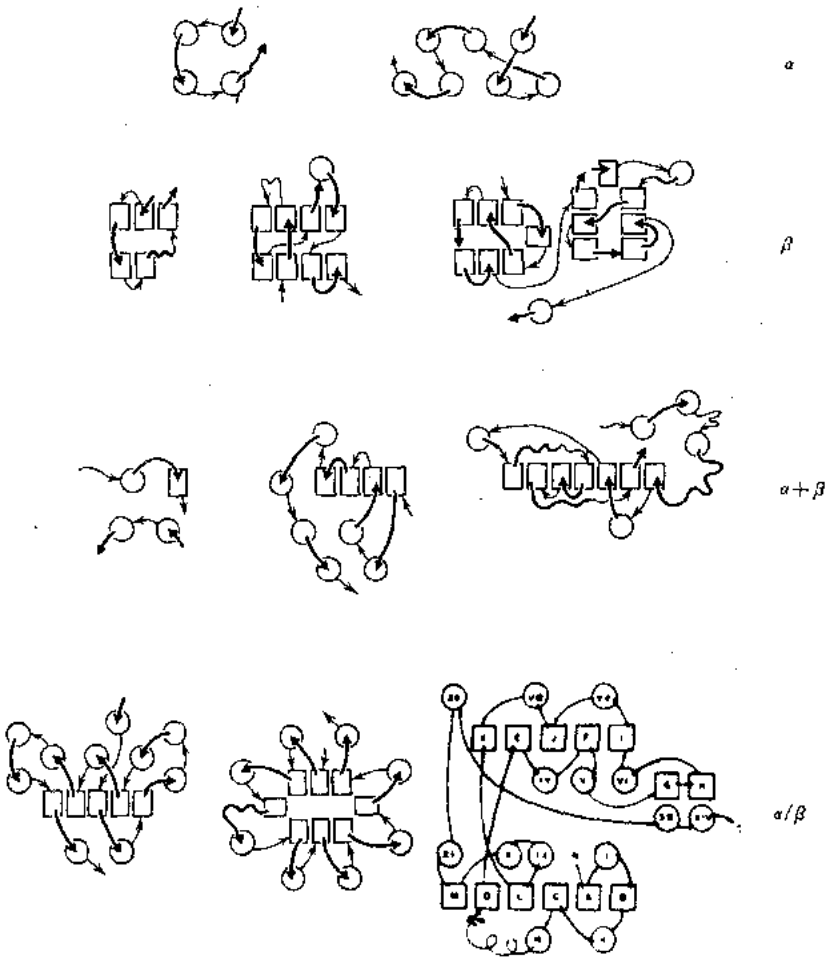


图7.57　球状蛋白质按三级结构分类的4种主要类型的实例拓扑图

○代表α螺旋，□代表β折叠链。图中从左至右，从上到下分别为血红质，肌红蛋白，红素氧还蛋白，前清蛋白，胰凝乳蛋白酶，胰岛素，溶菌酶，胃蛋白酶，腺苷酸激酶，磷酸丙糖异构酶，麦芽糖结合蛋白

他部分，因而产生了某些生物效应如对各种配位体具有不同的结合常数等。当然，不相同的亚基之间的接触，则更容易是不对称的，例如在血红蛋白中(图7.62)，两个β亚基之间的接触比两个α亚基之间的接触较宽松些并形成了对若干种重要的效应分子的结合位置。另一种类型的非等连系存在于20面体的球形病毒中。例如番茄丛矮病毒(见第九章图9.17)，它仅具有60个对称等效位置，但却具有多于60个的亚基，即具有180个亚基。这种矛盾由亚基的特殊排列方式得以解决，即5个亚基围绕着每一个五重轴而相互连系，而6个其他的亚基围绕着每一个三重轴以与前者不同的但是相似的接触而相互连系。这种非等的连系的通用性使得更大的和更为复杂的一些生物学结构的组装成为可能，并且甚至更为普遍。

7.4.4.2　四级结构与生物学功能的关系

四级结构在从各个亚基组装成较大的生物分子或结构中起着主要的作用。而这些亚基聚合物的形成具有异常重要的生物学功能。特别是，亚基间接触的相互作用提供了各亚基之间的通讯方式，使得从一个配位体或底物分子与一个亚基的相互作用所产生的效应能够影响在这

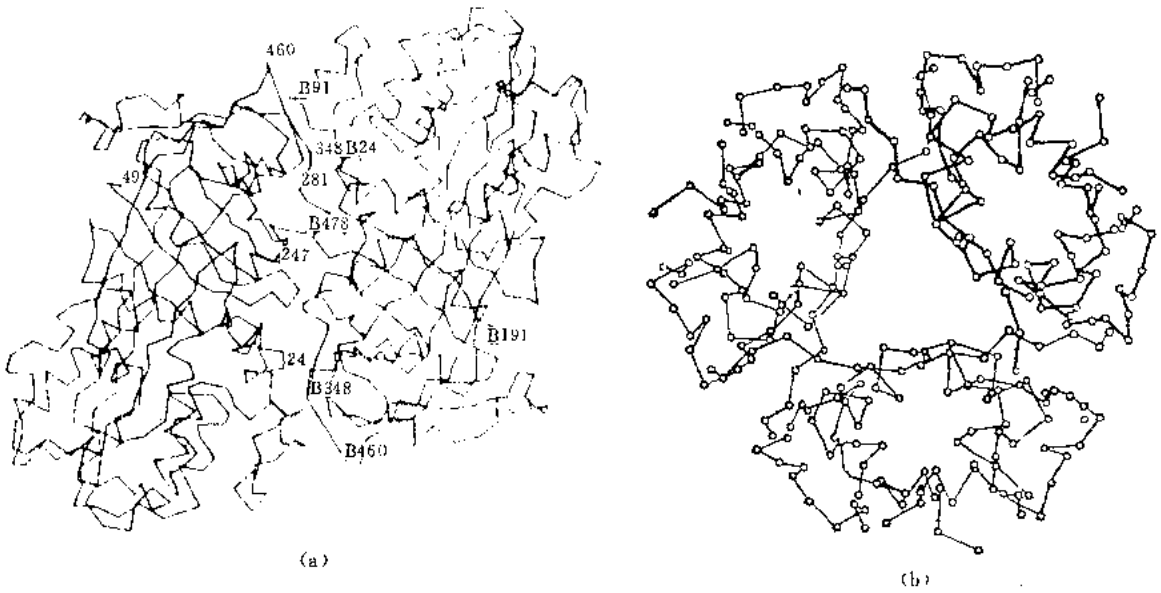


图7.58　规则排列的四级结构实例

(a)去六肽卵清蛋白二聚体;(b)胰岛素六聚体

个聚合体中的其他亚基随后形成的效应过程。这些相互作用形成了在生物化学系统中的协同效应的基础,为生物化学过程提供了许多调控机制。在这方面,X 射线晶体学分析方法发挥了最强大的作用,最突出的例子是 M. F. Perutz 等人[54,55,56,57]对血红蛋白的携氧功能与四级结构的关系所进行的开拓性研究。

血红蛋白是红血球的主要成分,其功能是运输 O_2。分子量约为64000,约为肌红蛋白分子量的4倍。它由两条相同的 α 链(α 亚基)和两条相同的 β 链(β 亚基)组成,每个亚基各含一个血红素,其中各有一个 Fe^{2+} 离子。每个亚基中有长短不一的8个螺旋区依次称为 A,B,C,D,E,F,G,H,盘绕成约为4.3nm×3.5nm×2.3nm 的扁球体。在分子表面具有一个疏水口袋,血红素在口袋中。Fe^{2+} 除了与血红素本身原卟啉的4个吡咯环的 N 原子配位以外,还与珠蛋白的螺旋 F_8 残基 His 的咪唑基配位,它的第6个配位体是可逆结合的 O_2,见图7.59。血红蛋白分子的4个亚基按近似于四面体的方式排列成四级结构,具有近似的222对称性,见图7.62中的示意图。

血红蛋白的 α 亚基和 β 亚基的三级结构都与肌红蛋白分子基本相同,但是血红蛋白分子与肌红蛋白分子的氧结合曲线却不同,见图7.60。氧合曲线表明,血红蛋白与 O_2 的结合量越大,则亲和力越强,即结合了 O_2 的亚基协助和促进未带氧的亚基与 O_2 结合。这种现象称为协同效应,许多蛋白质都表现这种效应。精确测定的血红蛋白的晶体结构,对血红蛋白的协同效应的分子基础——别构作用(或变构作用)提供了十分详细的模型。所谓别构作用是指当蛋白质分子的活性部位之一结合了其他分子如底物等时,随之而发生的蛋白质分子的构象、形状和性质的改变。以各种血红蛋白(如氧合的、去氧的、正铁的等等)晶体进行的 X 射线衍射结构分析,观察到下述一系列结构变化,并揭示了血红蛋白的协同效应是因为它具有两种不同的四级结构,即与配位体亲和力高的所谓松弛态 R 和亲和力低的所谓紧张态 T 两种构象。

在去氧血红蛋白晶体结构中观察到,铁原子的直径比卟啉环中央的孔大,因而铁原子不能

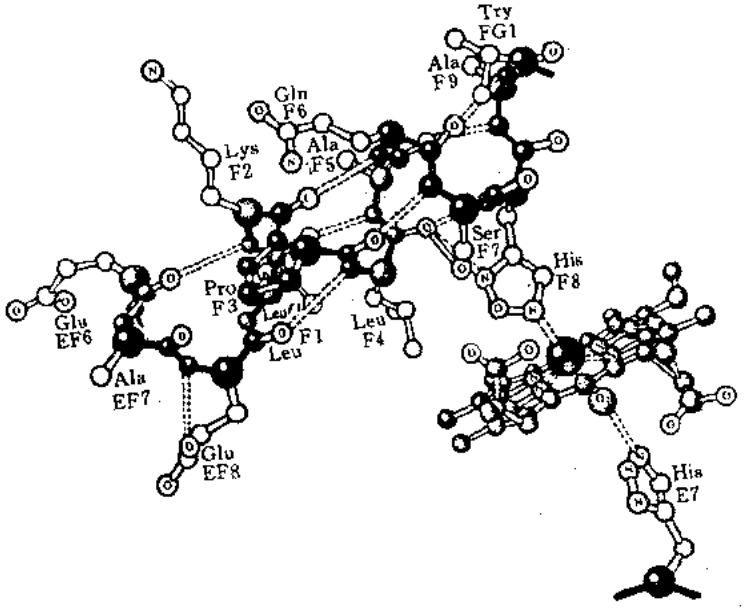


图7.59　氧合血红蛋白中 Fe^{2+} 的配位情况

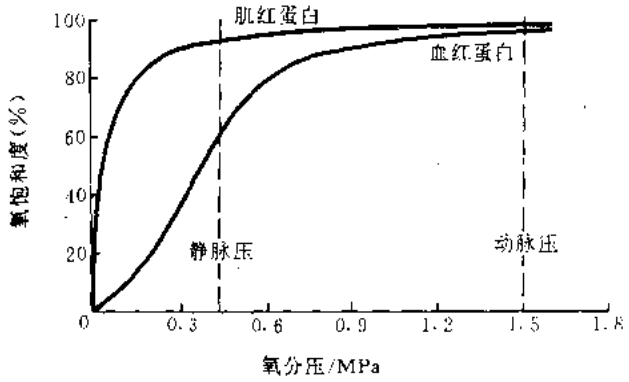


图7.60　血红蛋白和肌红蛋白的氧合曲线

落入中央孔中，而与血红素分子平面保持0.075nm 的距离，见图7.61。α 亚基由于 O_2 结合部位没有空间障碍，因而能首先与 O_2 结合。当一个或两个α亚基与 O_2 结合后，使 Fe 原子直径缩小，落入卟啉环的中央孔中。铁原子的移位，牵动了与之配位的 F_8 残基 His，使之靠近血红素；而 F_8His 的移动又引起亚基构象的相当大的变化，螺旋 F 向螺旋 H 靠近，使疏水性口袋变窄并将原来在口袋中的 HC_2 残基 Tyr 挤到口袋之外。HC_2Tyr 是 C 末端倒数第二个残基，因此，这一变化使 C 末端肽段游离出来，从而破坏了去氧血红蛋白分子中亚基间形成的某些盐键，导致四级结构构象的变化。亚基的重排使两个 β 亚基靠近了，见图7.62。同时，β 亚基的构象也发生某些变化，从而排除了螺旋 E_{11} 残基 Val 侧链对 O_2 结合部位的空间障碍，使得 β 亚基也能够与 O_2 结合。根据上述观察到的构象变化可以初步解释协同效应的现象：去氧血红蛋白分子由于

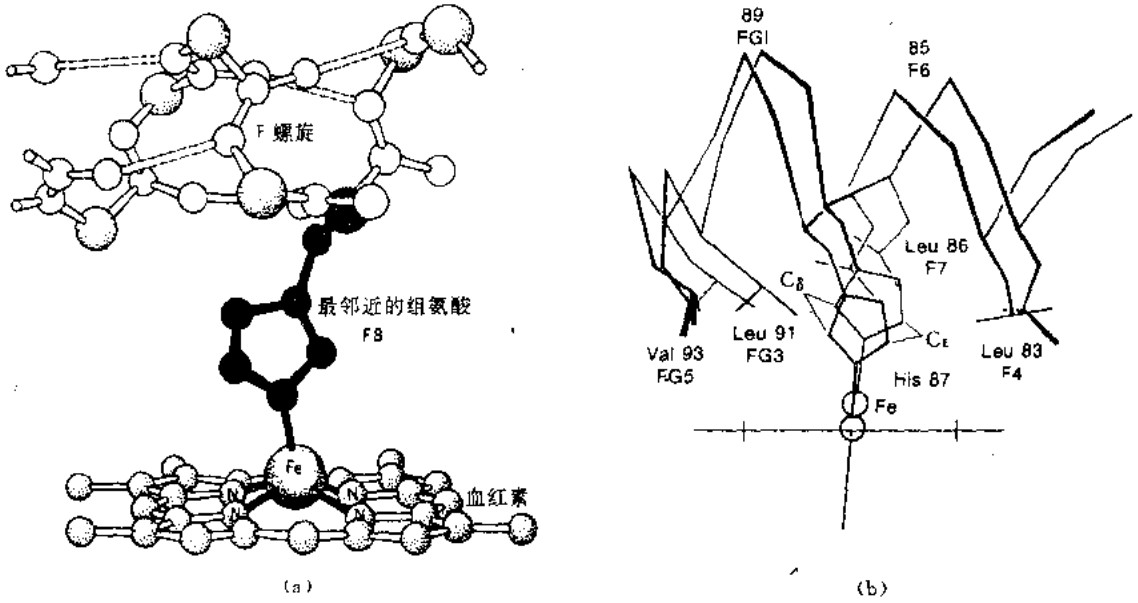
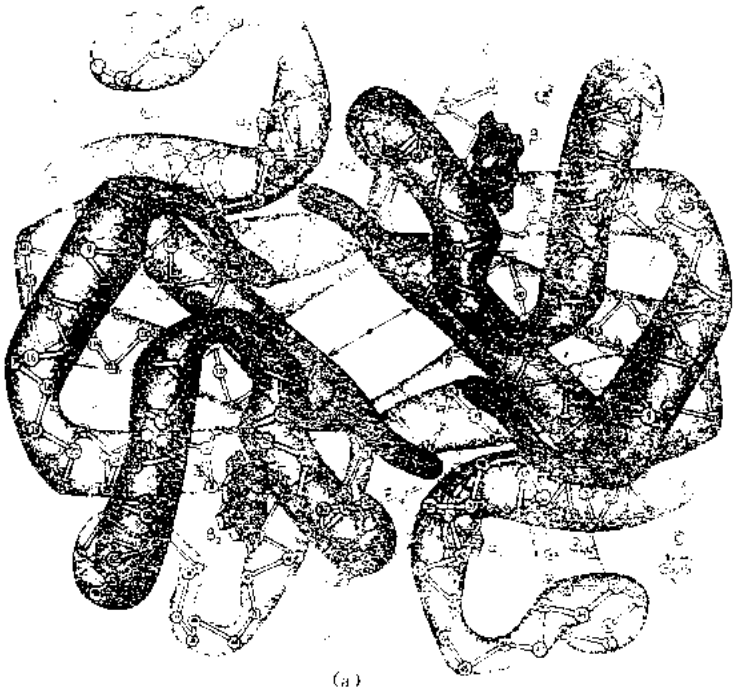


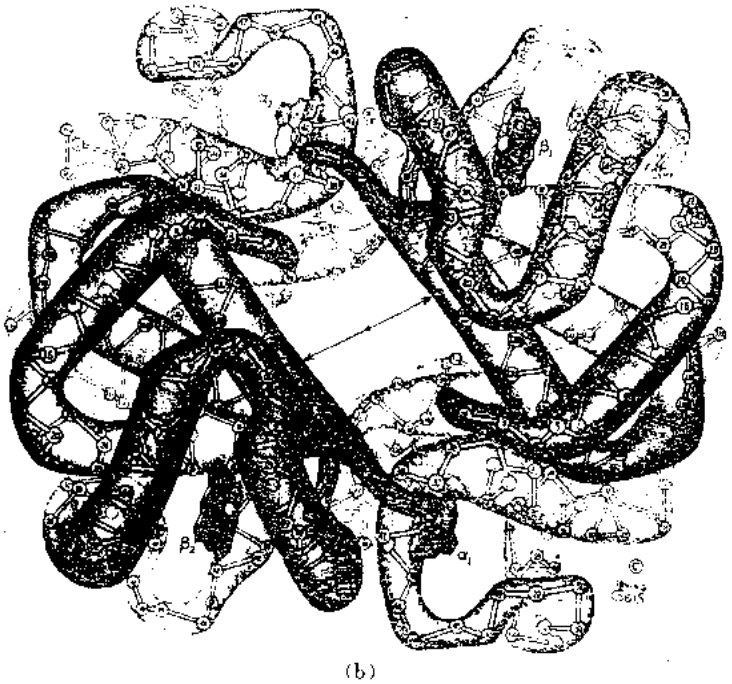
图7.61　去氧血红蛋白分子中的铁原子位置(a)以及氧合血红蛋白与去氧血红蛋白的血红素和F螺旋区域的叠合比较(b)

亚基间形成盐键，成为所谓的紧张态构象T，不易与O_2结合，只有某些盐键被破坏后，才能与O_2进一步结合，整个分子成为松弛态构象R。很可能当结合第一个O_2时，要打开4个亚基间的盐键，因而所需的能量多于结合第二个、第三个O_2时所需的能量，当结合第四个O_2时，所需能量更少。因此在氧合曲线上表现为，开始时随着氧分压增加，O_2结合量增加较慢；从结合了第一个O_2开始，则随着氧分压的增加，O_2结合量增加相当快，致使氧合曲线呈S型，而不同于肌红蛋白的双曲线形，见图7.60。

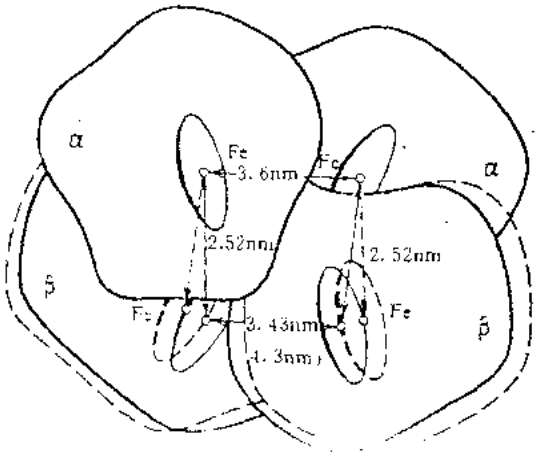
晶体学家还研究了二磷酸甘油酸(DPG)与血红蛋白的结合。DPG是血球中正常的化学成分，它可与血红蛋白分子($\alpha_2\beta_2$)以等摩尔数相结合。结合后，血红蛋白结合O_2的能力大大降低，氧合曲线向右移。晶体结构分析表明，DPG位于4个亚基之间的空穴中，正好填满原有的空隙，与两个β亚基的EF6残基Lys和H_{21}残基His的α氨基的正电荷相作用，见图7.63。因此，当去氧血红蛋白结合了DPG时，除了原有使紧张态稳定的一些作用力以外，又增加了DPG与β亚基形成的几对盐键，因此对维持紧张态有利而不利于与O_2结合；反之，当血红蛋白与O_2结合后，两个β亚基的螺旋H相互靠近，空隙变小，排斥DPG，降低了血红蛋白结合DPG的能力。因此，血红蛋白分子与O_2或与DPG的结合两者是相互排斥的。当血液流过氧分压较低的组织时，DPG与血红蛋白结合，使血红蛋白与O_2结合的能力降低，因而使血红蛋白所携带的O_2更多地被释放出来，因而有利于氧的输送。这一类影响蛋白质变构作用的因子称为变构效应因子，它们的存在使生物体又增加了一种调控因子，使有机体更好地适应环境的变化。



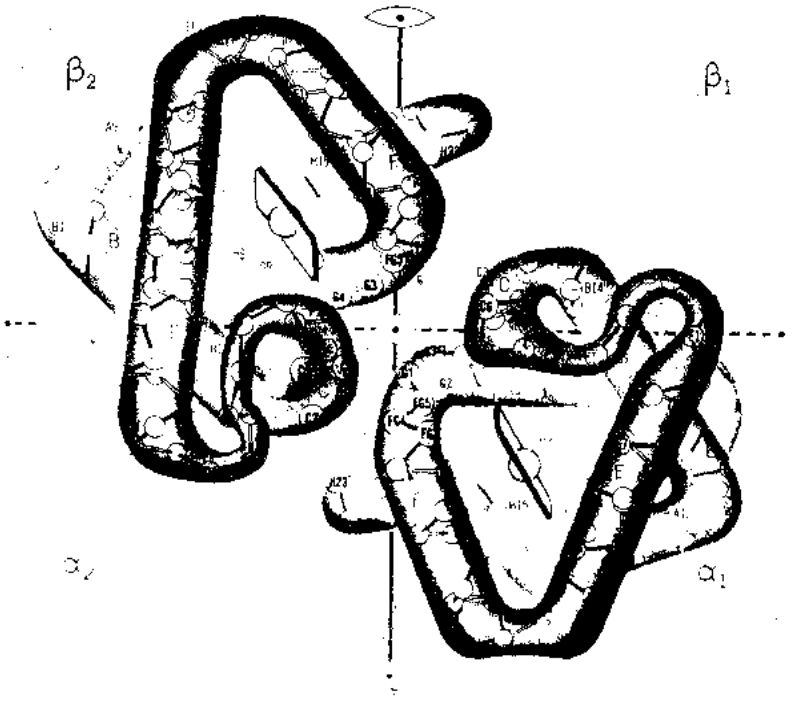
(a)



(b)



(c)



(d)

图7.62　血红蛋白四级结构的变构效应

(a)氧合血红蛋白;(b)去氧血红蛋白;(c)示意协同效应;(d)四聚体分子的非晶体学对称性示意图

(其中真二重轴由实线代表,假二重轴由虚线代表)

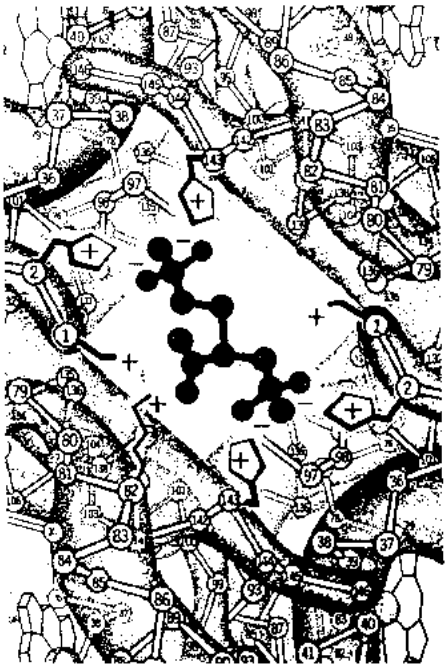


图7.63　血红蛋白与 DPG 的结合

7.5　膜蛋白的晶体生长和晶体结构

膜蛋白是非常重要的一类蛋白质，它是生物膜的主要组成成分，而且它的分离、纯化、结晶以及三维结构规律等都有别于前述纤维状蛋白质和可溶性球状蛋白质。因此，我们将在本节中专门进行介绍。

生物膜执行着生命过程中必不可少的重要功能。从低等生物直至高等生物，生物膜使细胞与环境隔离；真核生物细胞中还含有许多内膜，形成像线粒体、叶绿体和溶酶体等细胞器的界面；除此以外，生物膜也控制细胞与其环境之间的信息交流如接受信号、传递信号或发出信号等，其中的膜蛋白执行着一些独特的生物功能，如泵、门、受体、能量转换器和酶等。

不同来源或不同类型的生物膜的结构和性质差别是很大的，甚至于同一个细胞内的膜系统也是很不相同的。但是尽管如此，所有生物膜仍具有共同的基本结构特征：它们是片状的内外两侧不对称的结构；它们是主要由脂类和蛋白质（还有碳水化合物等）组成的非共价的闭合的双脂层，而且为流体结构，脂类分子和蛋白质分子除非为特定的相互作用所固定外，在膜平面内可以扩散、流动，因而可以认为生物膜是定向的蛋白质和脂类的二维溶液。图7.64为生物膜的结构示意图。

膜脂类是较小的分子，主要有三种类型，即磷脂、糖脂和胆固醇。虽然种类很多，但它们具有共同的结构特点，即膜脂类均为两性分子，包含亲水部分（又称为极性头基团，一般以圆圈代表）和疏水部分（又称为烃尾，一般用直线或波线表示）。图7.65中例举了两个膜磷脂分子的结构式和它们的空间填充模型。它们的极性头基团对水有亲和力，而烃尾要避开水。因此，这些膜

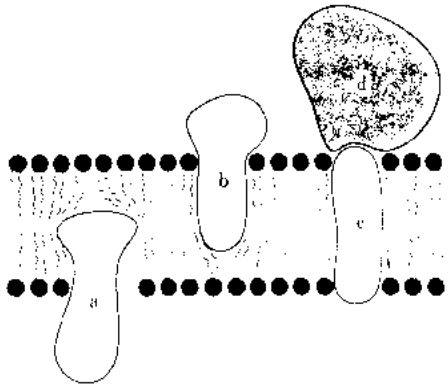


图7.64　生物膜的结构示意图

a,b,c 代表膜本体蛋白,它们与双脂层的烃区有广泛的相互作用;

d 代表膜周围蛋白质,它们与膜本体蛋白的表面相结合

脂在水介质中有两种可能的排列方式能满足这一要求。一种排列方式是形成胶粒,另一种方式是形成双分子层片,见图7.66。实际上大多倾向于自动地形成脂双层片状结构。疏水相互作用是形成这种脂类双层的主要推动力。烃尾之间的范德华吸引力也有利于烃尾的密堆积;极性头基团与水分子之间则有静电和氢键相互作用。

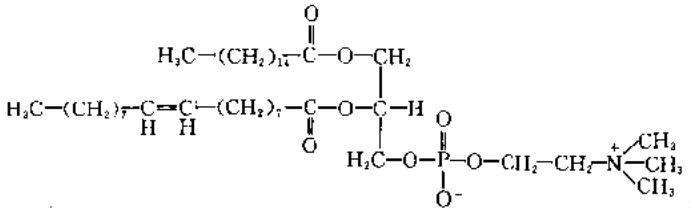
膜蛋白根据在膜上的位置和分离的难易可分为周围的和本体的两类。本体膜蛋白与膜脂的烃尾有广泛的相互作用,因此要将膜蛋白提取分离出来只有用能与这些非极性相互作用竞争的试剂如各种去垢剂或其他有机溶剂才能实现;而周围膜蛋白则只是通过静电和氢键结合在膜的外表面上,因此只要用加入盐类例如 1mol/lNaCl 或改变 pH 值的方法等就能破坏这些键而将此类膜蛋白分离出来。大多数周围膜蛋白是结合在本体膜蛋白的表面上的。

7.5.1　膜蛋白的晶体生长[58,59]

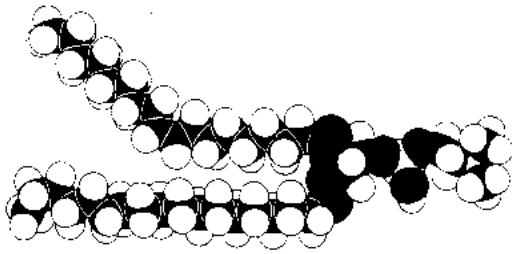
对于纯化了的可溶性球状蛋白质来说,经过几十年的研究,要得到适合于 X 射线衍射分析的单晶已不是十分困难的了。但是纯化了的膜蛋白曾经是、现在仍然是异常难以结晶的。这是由于它在膜内部的蛋白质部分是疏水的,暴露在膜的内外表面的蛋白质部分是亲水的,这种亲脂又亲水的结构使得它们一旦从天然的脂环境中提取出来,其非极性区域则倾向于形成无序化的聚集物。因此,不能用一般可溶性蛋白质的结晶规律和条件,而必须采用特殊的条件才能使膜蛋白形成晶体。

我们知道,膜蛋白在水溶液中要有去垢剂存在时才能溶解,但除掉了去垢剂以后蛋白质就会重新聚合。去垢剂本身也有像膜脂分子一样的两亲性结构,在相当高的浓度时,去垢剂可以形成胶粒。在胶粒里,去垢剂分子的烷烃链就被包围在胶粒内,使它不能与水接触。当用去垢剂溶解膜蛋白时,实际上就是用去垢剂胶粒去替换原来的双脂层。这样,去垢剂胶粒的烷烃链就与蛋白质的疏水部分结合,去垢剂的亲水极性头基团就与蛋白质的极性表面结合。图7.67为用去垢剂溶解蛋白质的示意图。

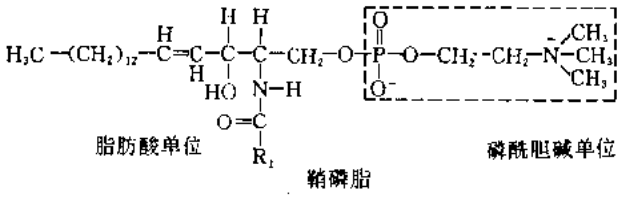
因此,从膜蛋白的水溶液中形成膜蛋白晶体有两种可能的类型。这两种可能情况中所不同的是用不同的方法来替换或补偿蛋白质疏水部分原来所结合的膜脂,见图7.68中的示意图。图中 I 型是由几个相似的膜蛋白分子组成的二维晶体片层再堆积成三维晶体。在此,蛋白质的疏



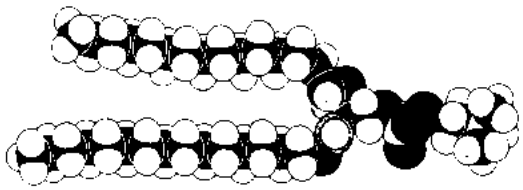
磷脂酰胆碱



(a)



鞘磷脂



(b)

图7.65　膜脂分子的结构举例

(a)磷脂酰胆碱的结构式和空间填充模型；(b)鞘磷脂的结构式和空间填充模型

水部分是用脂质分子去结合的，所以二维片层只有极性部分暴露在外，二维片层之间靠极性吸引力形成三维晶体。对于若干种膜蛋白，已用这种方法得到了这类晶体。但是这类晶体一般很小，不能用于高分辨率衍射的研究。图中Ⅱ型晶体是用如图7.68中所表示的由胶粒包围的膜蛋白所形成的。在此，晶体主要是靠膜蛋白之间的极性吸引力去稳定。由图可看出，显然膜蛋白分子不能太小，否则蛋白质极性区域的接触不够。可是，蛋白质分子的大小是无法改变的。因此，外围的去垢剂的大小和电荷以及其他化学性质就成为关键的了。也就是说，要选用适当小一点的两亲性分子去置换去垢剂分子，使其能基本上填满于空隙中。不合适的去垢剂的相互作用会使晶体不稳定。

在膜蛋白的晶体生长研究方面，最卓越的是 H. Michel[41,59]。他于1980年曾用Ⅰ型方法获得了细菌视紫红质的晶体[60]，于1982年[61]，他用Ⅱ型方法，选用庚烷-1，2，3-三醇来代替去垢剂，首次成功地得到了膜蛋白绿色红假单胞菌光合作用反应中心的大晶体，这些晶体的线度可

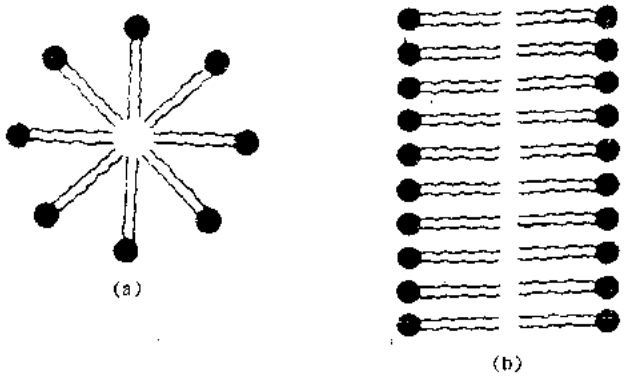


图7.66　膜脂分子在水介质中的两种可能的排列方式

达0.4×0.4×1mm，而且还保留了反应中心的光学活性。又于1984年[41]和1985年[62]连续发表了0.3nm 分辨率 X 射线衍射分析的结果。由于这一工作对于阐明光合作用中心的膜蛋白三维结构的重大贡献，H. Michel，J. Deisenhofer 和 R. Hüber 三人共同获得了1988年诺贝尔化学奖。我们就以这一工作的结晶条件为例。

从光合细菌绿色红假单胞菌分离纯化出来的光合作用反应中心，含有4个不同的蛋白质亚基，分别称为 H（重的意思，分子量为35000道尔顿），M（中等的意思，分子量为28000），L（轻的意思，分子量为24000）以及一个结合在膜上的 c 型细胞色素。结晶时，仍选用分离纯化时所用的去垢剂 N，N，十二烷基二甲胺 N-氧化物*（DDAO）。因为这个去垢剂的大小和极性与 β-辛烷基-吡喃葡萄糖苷相似，而后者曾成功地用作另外两个膜蛋白结晶作用中的去垢剂，即 Michel 本人[61]在1980年用于细菌视紫红质的结晶作用以及 M. Garavito 等人[63]在1980年用于大肠杆菌外膜骨架微孔蛋白（OmpF 蛋白）的结晶作用。但是，在绿色红假单胞菌的光合作用反应中心的结晶实验中，只有当同时还存在有小的两亲分子如庚烷-1，2，3-三醇时，才能得到晶体。晶体生长的具体条件为[62]：

将分离纯化的最后一步柱层析所得到的含有

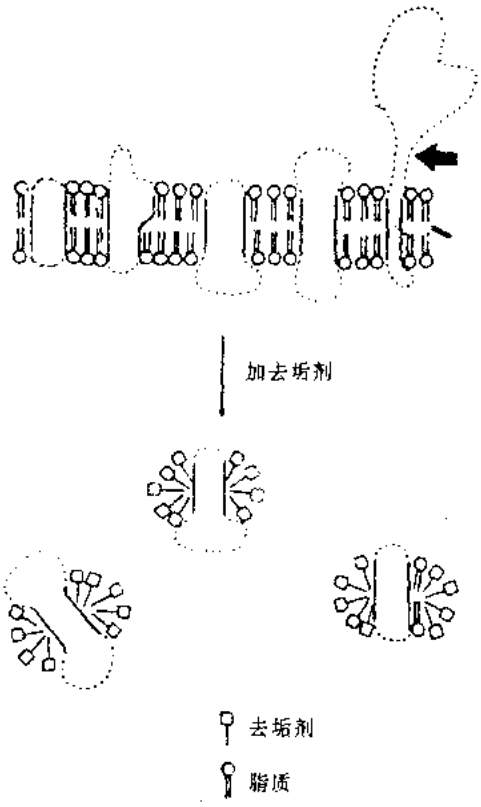
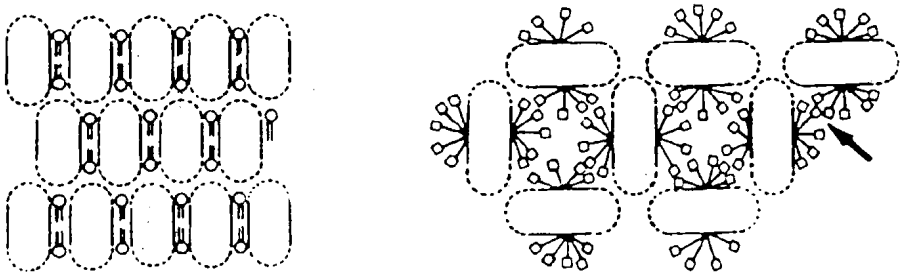


图7.67　用去垢剂溶解膜蛋白的示意图
图上部为双脂层与蛋白质组成的生物膜。虚线代表膜蛋白的亲水表面，实线代表膜蛋白的疏水表面。圆头双尾代表膜脂分子，方头单尾代表去垢剂分子

* DDAO 的结构式为 $H_3C-(CH_2)_{10}-CH_2-N^+(CH_3)_2-O^-$。



Ⅰ型　　Ⅱ型

图7.68　膜蛋白结晶的两种主要方法

反应中心的各收集部分合并，此时的洗脱缓冲溶液含有20mmol/l 磷酸钠盐(pH7)、0.1%(*W*/*V*)，N,N,十二烷基二甲胺 N-氧化物以及0.1%(*W*/*V*)叠氮化钠。然后置于干燥器中以五氧化二磷为干燥剂浓缩6倍。再加入至最终浓度为3%(*W*/*V*)的庚烷-1,2,3-三醇、3%三乙胺硫酸盐以及1.5mol/l 硫酸铵，并调节 pH 值至6。离心去除变性了的蛋白质，然后用汽相扩散法生长晶体。取上述含有反应中心(在830nm 下吸收值为2个单位)的溶液在室温下对2.4mol/l 硫酸铵溶液进行汽相平衡。在3周之内得到了大的暗褐色四方柱形晶体，线度可达0.4×0.4×1mm，衍射分辨率达0.25nm 以上。空间群为 $P4_12_12$或其对映体 $P4_32_12$，晶胞线度为22.3×22.3×11.4nm。图7.69和图7.70分别为晶体照片和它的 X 射线旋转照相法衍射照片。

图7.69　绿色红假单胞菌光合作用反应中心的晶体

至今已有若干种本体膜蛋白被结晶了。虽然成功的数目是很少的，而且不是所有这些得到的晶体都适合于 X 射线衍射结构分析，但是我们已经知道，本体膜蛋白是能被结晶的，而且在结晶作用的方法学上已总结出几条一般原则：(1)本体膜蛋白是用去垢剂被结晶的；(2)去垢剂的选择是重要的，即要选择合适的两亲性离子或非离子的去垢剂；(3)小的两亲性有机分子能有助于结晶作用，这很可能是由于它们影响了去垢剂极性头基团之间的相互作用；(4)通常已成功地用于可溶性蛋白质结晶作用的沉淀剂如聚乙二醇和硫酸铵，也曾成功地用于膜蛋白以诱发结晶作用。

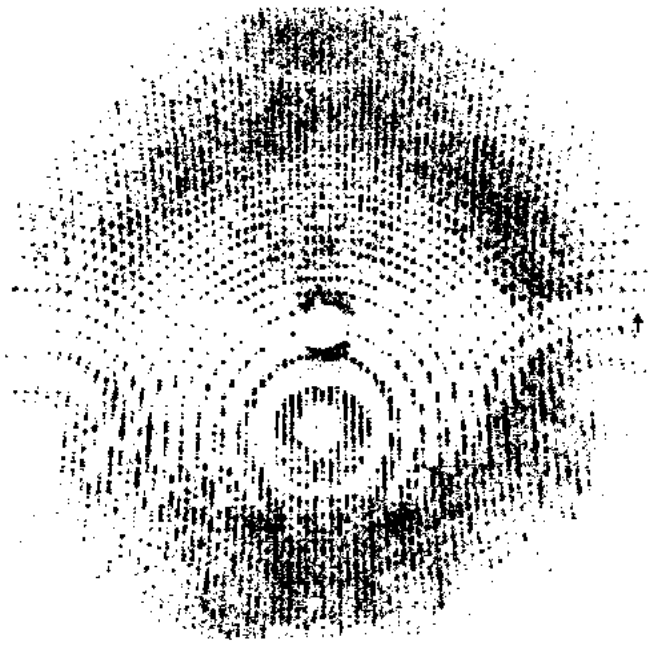


图7.70 绿色红假单胞菌光合作用反应中心晶体的旋转法衍射照片

7.5.2 膜蛋白晶体的X射线衍射分析方法

与可溶性球状蛋白质一样，膜蛋白晶体的结构分析也主要用重原子同晶置换法和分子置换法等。例如在绿色红假单胞菌光合作用反应中心的结构分析中[41]，找到了若干个不扰乱晶体有序性的重原子试剂，得到5种重原子同晶型衍生物，用于结构测定，它们是MDDB(2-甲基-1,4-二氯汞-2,3-二羟丁烷)，TAMM(四乙酰氧基汞甲烷)，URNI(硝酸铀)，DAME(乙基-二乙酰氧基汞氯代乙酸盐)以及PCMS(对-氯汞苯磺酸盐)。用同晶置换法解出了结构。又例如球形红假单胞菌的光合作用反应中心复合物晶体结构[64]则是在此后用分子置换法测定的。

除此以外，还有人用高角度漫射X射线衍射法，粗略推测膜蛋白的二级结构及它们在膜中的排列方向。例如，大肠杆菌外膜上形成通道的微孔蛋白[65]。将其晶体以合适的方向拍摄静止照片，见图7.71。根据衍射花样中在子午线方向和赤道线方向上最强的若干个衍射斑点所代表的重复距离，来推测其二级结构以及在膜中的排列方向。从图中可以看到：在子午线方向上有0.35nm的成双的反射(相当于沿β股的重复距离)、0.46nm反射(是β片层的另一个特征反射)以及0.57—0.62nm成双的反射(尚未被解释，可能是由于结合在膜蛋白上的去垢剂所致)，而无0.51nm反射(α螺旋的特征反射)；在赤道方向上，有代表β股和片层堆积的0.46nm反射和0.95nm反射。其中0.95nm反射表明为反平行β股之间的重复距离。上述这些参数与β丝蛋白和β角蛋白的衍射花样相同，但不同于可溶性球状蛋白质中的反平行β折叠片层，因为后者很少为理想的折叠片层。从而可以判断，这个膜蛋白是以β片层占优势的，而且β片层基本上垂直于膜平面。由0.35nm双反射的宽度可以推测出β股的平均长度为3.8nm，又由于0.35nm相当于沿β股的平均周期，所以每个β股平均含有10—11个残基。不过，尚不能判断它们是β桶结构还是堆砌的结构。

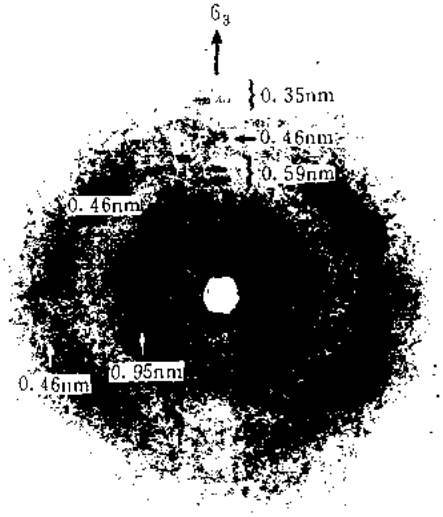


图7.71 大肠杆菌外膜微孔蛋白的一个六方晶体的漫射X射线衍射花样

7.5.3 膜蛋白的晶体结构

X射线衍射方法曾对分析可溶性蛋白质的三维空间结构发挥了巨大的威力。对于膜本体蛋白来说，则主要由于难以得到晶体，至今仍只有绿色红假单胞菌光合作用反应中心的晶体结构已被分析至高分辨率[41,62]，以及大肠杆菌的外膜骨架微孔蛋白已接近于解至高分辨率。目前，研究膜蛋白立体结构的方法除了X射线衍射法以外，还有电镜术三维重构方法和溶液拉曼光谱以及圆二色性方法等。

由于膜蛋白是在膜脂双层特殊的环境中，因此前面讨论过的关于可溶性蛋白质的结构规律对于膜蛋白来说则不一定适用。可以推测，膜蛋白的结合在膜中的片段与膜脂的接触必须是有序的构象，但是它们可能是以α螺旋或以β股的构象存在，这是由于膜蛋白在脂双层环境中必须满足分子内部的氢键，因为脂质的碳氢链不能参与氢键的形成。在α螺旋的情况下，沿着螺旋可形成链内氢键。因此，一个螺旋就可以代表一个在膜中的稳定的结构；而在β股的情况下，由于氢键只能在相邻的股之间形成链间的氢键，因此单一个β股在膜中不能作为一个稳定的结构，而是每一个β股必须至少有两个相邻的β股，这样才能形成一个紧密排列的由链间氢键结合成的稳定结构，即β桶。一般至少要求6—8个β股以形成一个β桶。

许多证据表明：在膜中，单个的螺旋常常作为暴露于细胞表面的蛋白质的一个固定点，例如红细胞的血型糖蛋白[66]和流感病毒的血细胞凝集素[67]；此外，单个的螺旋还可以作为允许其他颗粒通过膜而进入细胞的微孔。当然，对于一个每圈具有3.6个残基的α螺旋来说，作为这样的微孔是太小了，但是，一个每圈具有4.4个残基的π螺旋，则是宽到足以允许至少是质子的进入，例如短杆菌肽[68]。要使较大的颗粒能进入，则要求若干个螺旋来形成一个圈，中间有一个孔。在最简单的情况下，这些螺旋可能都是相同的，例如melittin[69]微孔可能是由4个单体形成的，又如alamethicin[70]则包括8—14个单体。具有更为复杂的功能的蛋白质，如泵蛋白质通常是由在同一个单肽链上的若干个不同的螺旋所组成的，例如细菌视紫红质[71]可能是由7个螺旋组成的，大肠杆菌的乳糖通道[72]含有10—14个螺旋，又例如绿色红假单胞菌光合作用反

应中心[41,62]含有分别在三个亚基上的共11个螺旋。

β桶类型的膜蛋白可能作为微孔，因为甚至于由6—8个β股组成的最小的β桶，其中心也具有直径为零点几纳米的孔，例如大肠杆菌等革兰氏阴性细菌的外膜中形成微孔的微孔蛋白等[65,73]。

以上例子大多是由电镜术三维重构方法和各种溶液构象测定方法研究所得到的结果。但是，正如我们所知，X 射线衍射法能给出结构细节。随着膜蛋白的晶体生长方法日趋成熟，将会有更多的膜蛋白精细结构被解析。

图7.72和图7.73为接近于原子分辨率水平的第一个完整膜蛋白的结构，即绿色红假单胞菌光合作用反应中心的0.3nm 分辨率 X 射线衍射分析所得的结构。这个反应中心复合物是长形的。图7.72(a)为反应中心复合物在膜中的位相和电子传递过程。图中虚线表示复合物的轮廓，它的4个亚基的位置分别以 L，M，H，C 标明；并绘出了复合物的所有辅基的原子模型，即4个细菌叶绿素 b，2个细菌脱镁叶绿素，1个非血红素铁，1个甲基萘醌以及4个血红素；图中并以箭头标明了电子传递的方向和过程。图7.73为亚基 L，M，H 以及 C 的结构。它们分别由273，323，258和336个氨基酸残基组成。

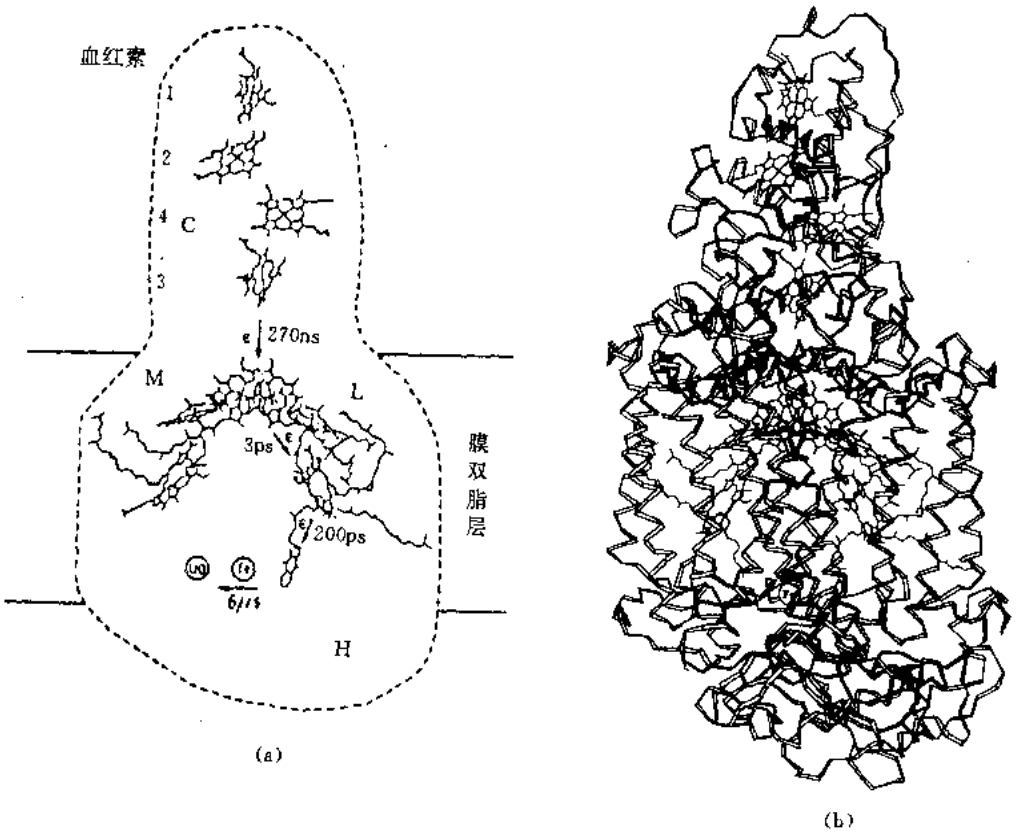


图7.72 绿色红假单胞菌光合作用反应中心复合物在膜内的位相和电子传递过程示意图(a)以及完整复合物的结构(b)

从图7.73可以看到，L 和 M 的空间结构大致相同，它们的多肽链的中部含有5段 α 螺旋、

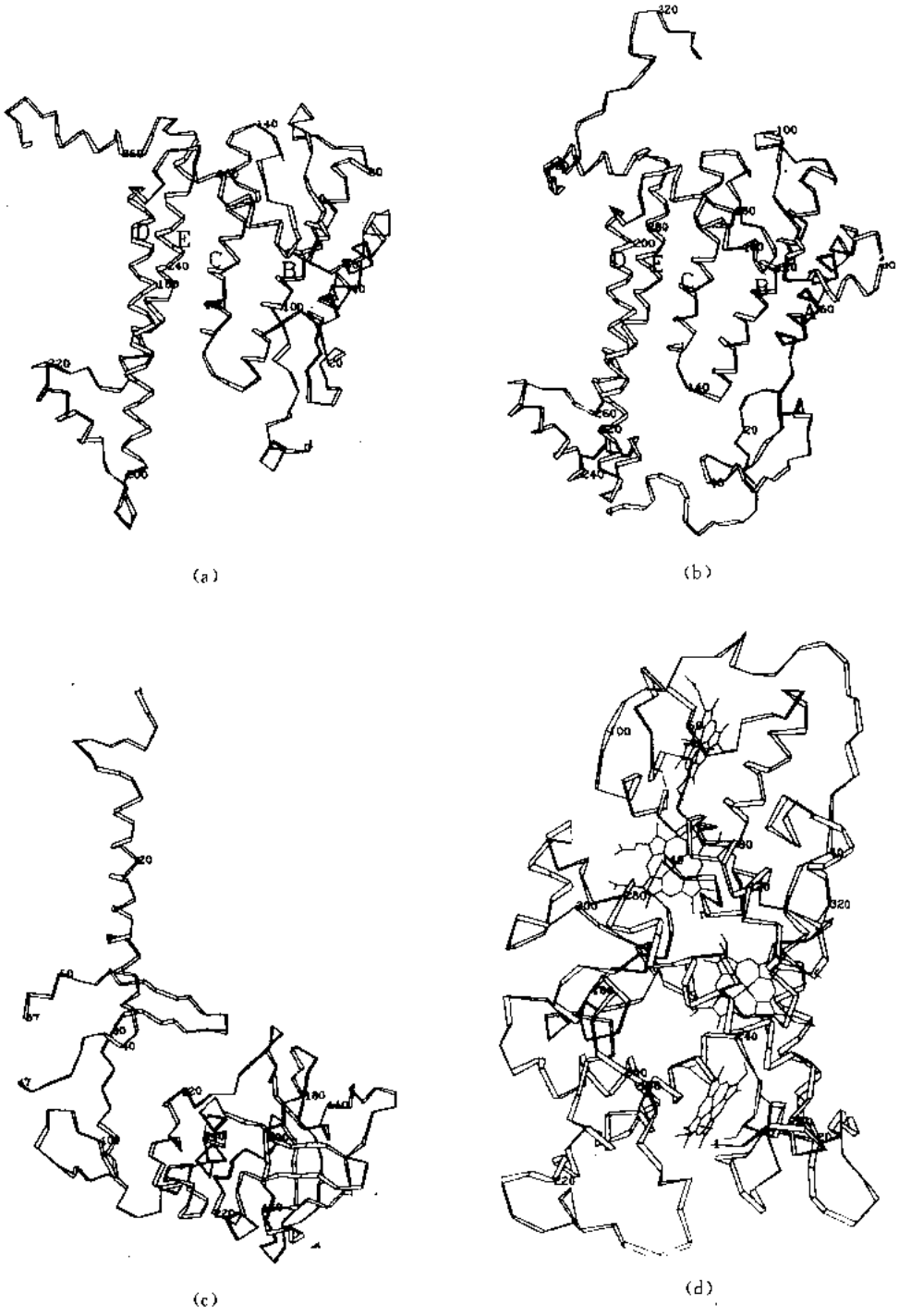


图7.73　绿色红假单胞菌光合作用反应中心复合物的4个亚基的结构
(a)L 亚基;(b)M 亚基;(c)H 亚基;(d)C 型细胞色素

分别称为 A,B,C,D,E。每段螺旋含有24—30个疏水性氨基酸。这5段螺旋中,A,B 和 D 三段是反向平行的,C 和 E 是平行的。这5段螺旋都具有跨膜的长度。螺旋区域的一端(图中上端)的 A 与 B 和 C 与 D 之间的短螺旋以及羧基末端是与 C 亚基紧密接触的;另一端(图中下端)的 B 与

C 和 D 与 E 之间的短螺旋以及氨基末端是与 H 亚基紧密接触的。H 亚基实际上是这三个亚基中最小的一个。它的氨基末端与 C 亚基接近；它的第12个至第37个氨基酸残基形成一段 α 螺旋，与 M 亚基的 D 和 E 螺旋相接触，也具有跨膜的长度。

细胞色素(C 亚基)位于外周胞质的一侧；H 亚基折叠成位于细胞质一侧的球形部分以及一个跨膜的 α 螺旋；L 和 M 亚基参入到磷脂双层中，它们的跨膜 α 螺旋被包埋在双脂层中。因此，由26个残基(H 亚基)和24—30个残基(L 和 M 亚基)所组成的共计11个长度合适的 α 螺旋一同跨过膜双脂层。这些片段的氨基酸序列除了在靠近 α 螺旋的两头末端处有很少几个带电荷的残基以外，没有带电荷的残基，而且这些残基的疏水性都很大。

L-M 复合物的净电荷分布是不对称的。在靠细胞质的一侧有6个净正电荷；而在靠外周胞质的一侧有8个净负电荷。这为复合物正确地插入膜双脂层中提供了较低的自由能障碍，因为细胞内膜电位是负的。H 亚基在跨膜 α 螺旋的 C 末端处也具有很强的极性的氨基酸序列，并有一段连续的带电荷残基的序列(H33—H39)，它可能有效地终止了插入膜中。在 L 和 M 亚基中，没有与此相似的序列，但是它们的 α 螺旋在细胞质一侧处的每一个连接片段中，都有3—11个带电荷的残基，它们也可能对 α 螺旋的正确定位起重要作用。

第八章　核酸的立体结构

核酸是生物体中另一类最重要的长链多聚物分子，它们是由核苷酸组成的。核酸包括核糖核酸(RNA)和脱氧核糖核酸(DNA)两大类，它们对生物体内遗传信息的传递及蛋白质的生物合成等生命过程起着决定性作用。因此，研究核酸的三维立体结构与功能的关系，对于揭示生命的本质是不可缺少的一个领域。目前，X射线衍射技术仍然是测定核酸立体结构最有效的方法。

8.1　核酸的结构和构象

RNA和DNA是化学上不同类型的核酸，两者均由各自的主要4种核苷酸所组成。每一个核苷酸是由一个核糖(RNA中)或脱氧核糖(DNA中)、一个与糖的C-1′相连接的含N碱基即嘌呤或嘧啶以及一个磷酰基团组成的。表8.1为DNA和RNA的化学组成比较，图8.1为RNA和DNA中单核苷酸的化学结构。

表8.1　DNA和RNA的化学组成

	戊糖	嘌呤碱	嘧啶碱	酸
RNA	β-D-核糖	腺嘌呤(A)，鸟嘌呤(G)	胞嘧啶(C)，尿嘧啶(U)	磷酸
DNA	β-D-2-脱氧核糖	腺嘌呤(A)，鸟嘌呤(G)	胞嘧啶(C)，胸腺嘧啶(T)	磷酸

若取具有5′-磷酰基的单核苷酸为结构单位，将一个这样的单核苷酸以磷酸酯键连接至另一个单核苷酸的3′-OH基团上，则形成一个二核苷酸。重复这种连接过程，则形成了多聚脱氧核糖核苷酸(DNA)或多聚核糖核苷酸(RNA)的一级结构。核酸的一级结构序列定义为从5′-磷酰基末端开始到3′-磷酰基末端方向的顺序。图8.2为一个DNA的一级结构示意图。RNA也具有类似的一级结构。

与蛋白质相似，核酸大分子中由于存在着可以自由旋转的单键，因而从理论上讲也可以形成无穷多种构象。首先需了解核酸中的构象角以及糖和碱基的可能的构象形式。

8.1.1　糖-磷酸主链和糖苷键的构象角

在核酸中，核糖或脱氧核糖与磷酸以3′，5′-磷酸二酯键相连形成核酸主链，并与碱基以C-1′-N-糖苷键相连形成侧链。因而在每一个核苷酸残基中有6个可以自由旋转的单键，其中5个在糖-磷酸主链中，1个在C-1′-N-糖苷键处，见图8.3[74]。这6个扭角的变化决定了核酸的构象。

8.1.2　核糖和脱氧核糖的4种折叠构象形式

核糖和脱氧核糖是五碳糖，它们在核酸中是以呋喃环的形式存在的。呋喃环可有4种折叠构象形式，见图8.4，它们在稳定性方面稍有差异。在核酸中，观察到了不同的折叠构象形式。

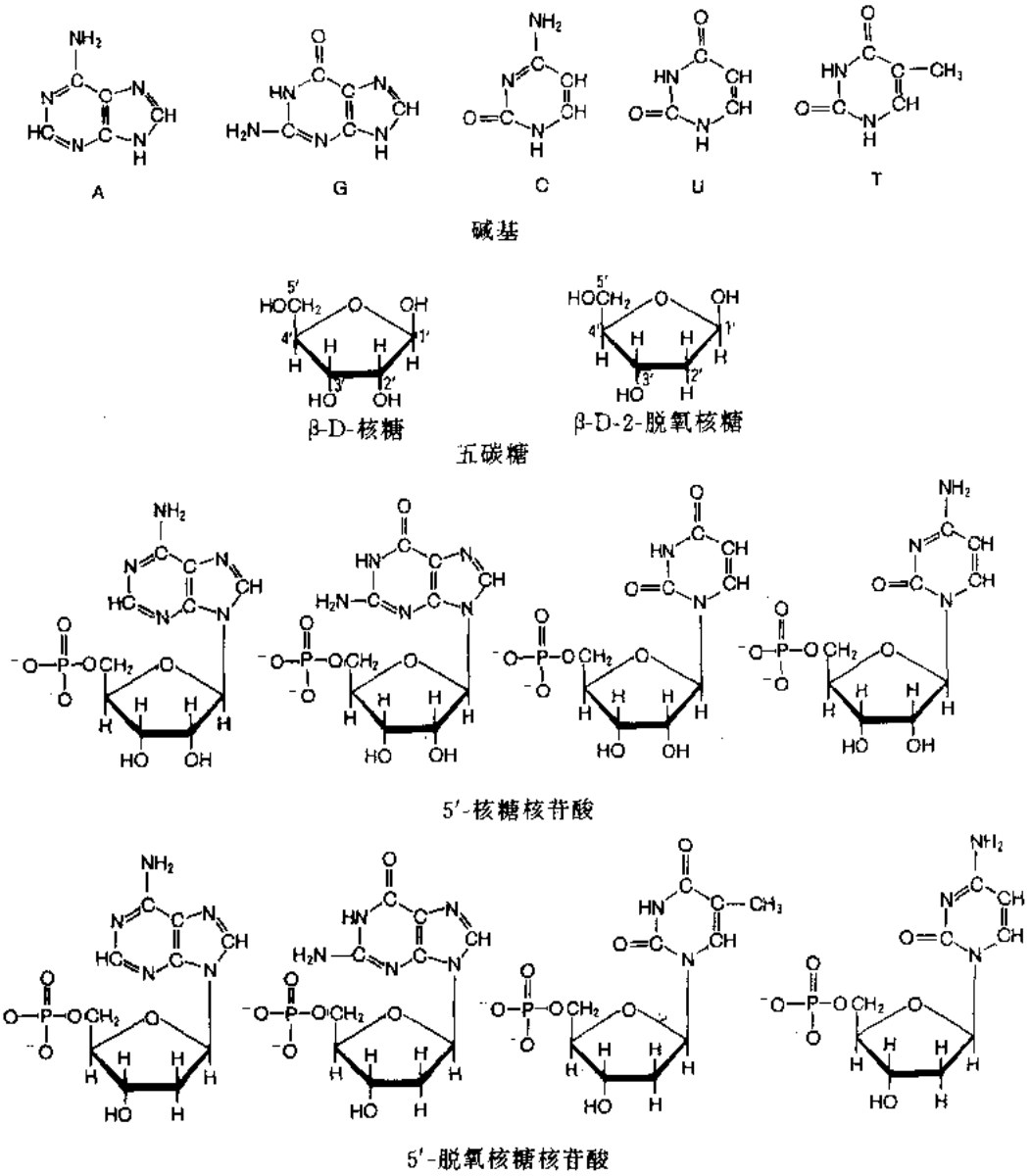


图8.1 RNA和DNA中常见的碱基、五碳糖及5′-单核苷酸的化学结构

8.1.3 碱基堆积的两种构象

由于C-1′-N-糖苷键的旋转，使得核酸中实际上存在着碱基相对于核苷酸的其他部分来说的所谓“顺式”和“反式”两种构象。模型研究表明，反式构象比顺式构象具有较小的空间障碍；嘌呤核苷酸比嘧啶核苷酸形成顺式构象较为容易。图8.5为这两种构象的结构。

8.2 脱氧核糖核酸(DNA)的立体结构

与蛋白质的立体结构研究工作一样，核酸立体结构的研究也是早在20世纪30年代由天然

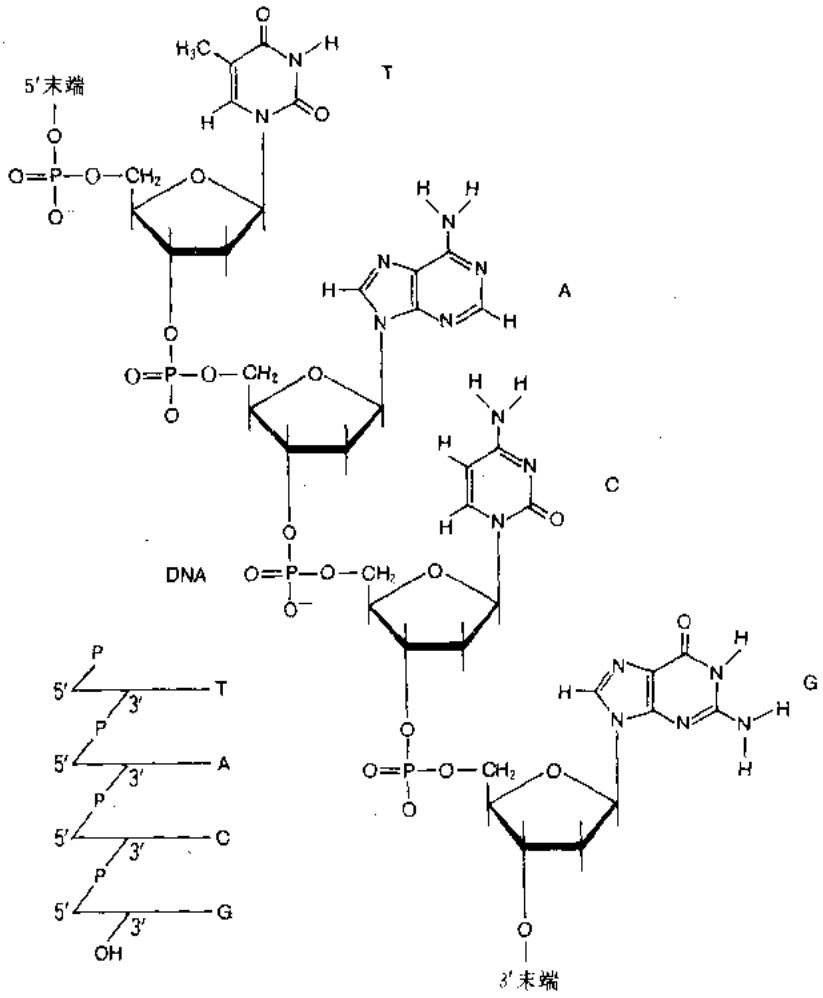


图8.2 DNA的一级结构示意图
图中结构应写作 pTpApCpG

DNA 纤维衍射研究开始的，并于1953年报道了由纤维衍射花样推测出的著名的 Watson-Crick 右手双螺旋 DNA 模型[76]，成为20世纪生物学史中具有划时代意义的成就。从70年代开始，DNA 结构的研究又进一步用人工合成寡核苷酸晶体进行单晶 X 射线衍射结构分析，证实了右手双螺旋 DNA 结构，并发现了左手双螺旋 DNA 结构的存在。DNA 的结构至今仍然是 X 射线晶体学研究的重要领域。

8.2.1 天然 DNA 纤维的制备和 X 射线纤维衍射花样

DNA 是以十分大的分子量（可从几万到几百万道尔顿）的长链形式存在于核蛋白中。它与蛋白质以较弱的结合力结合在一起，用温和的方法即可将它分离出来，并可使它以各种盐的形式沉淀下来。DNA 盐是纤维状的固体，它能摄取大量的水形成胶。用适当的机械处理和干燥方法，可以由这种胶拉成高度定向的 DNA 盐的纤维样品，用 X 射线纤维衍射照相法进行研究。图8.6为 DNA 纤维衍射原理图，与纤维状蛋白质纤维衍射相同。伸展的纤维样品中含有许多竖直的 DNA 分子链，一般说来，纤维样品的直径越小，纤维中 DNA 分子的定向程度越好。

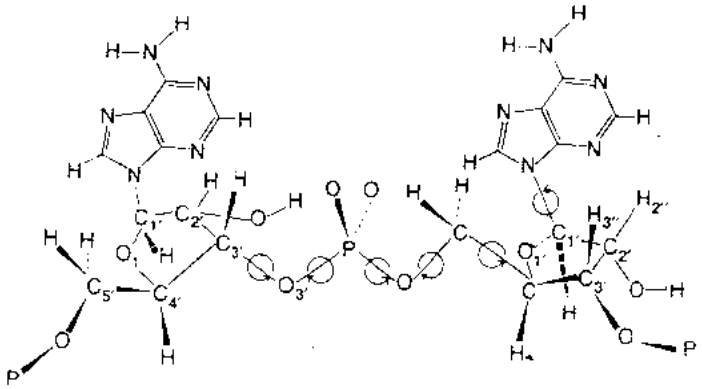


图8.3　核酸的构象角

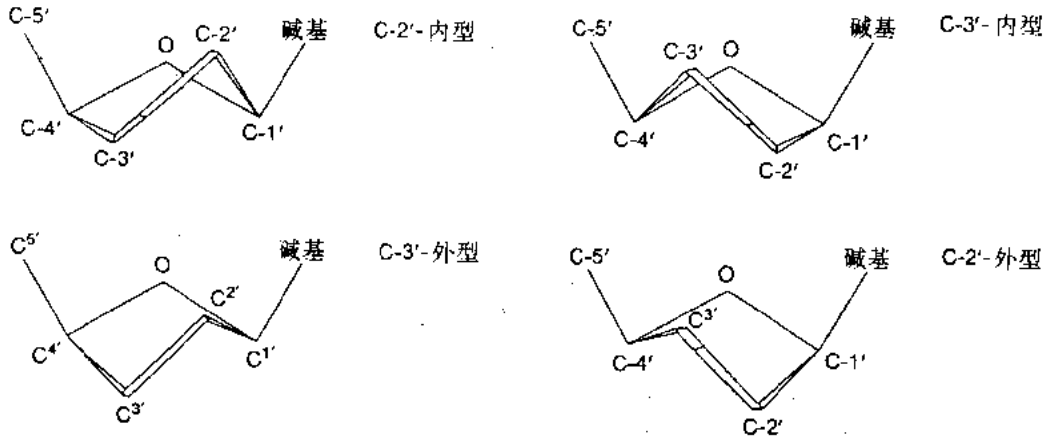
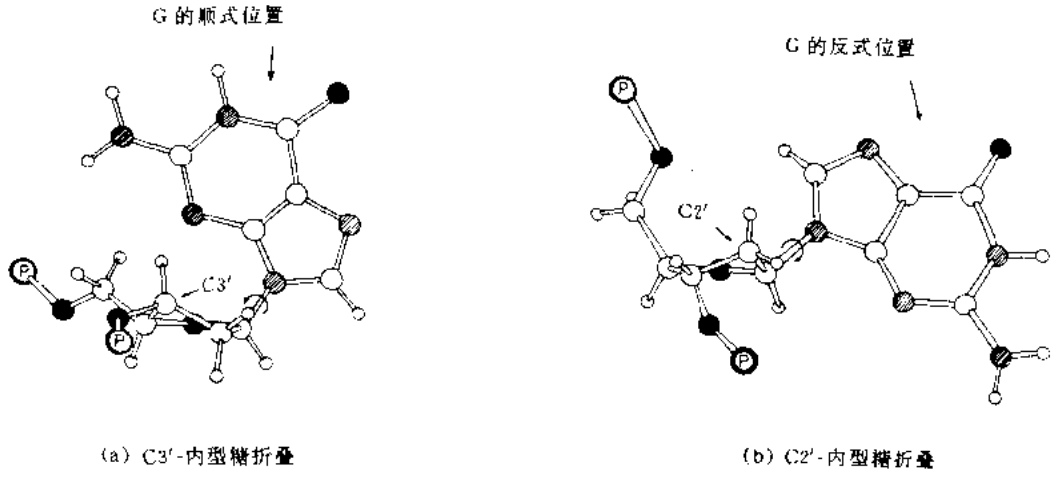


图8.4　核糖和脱氧核糖呋喃环的4种折叠构象形式

1938年报道了 W. T. Astbury[76]的开拓性工作。他用从小牛胸腺提取并制备的 NaDNA 纤维，拍摄到虽不十分好但可看到一个强的衍射弧的衍射照片。这个衍射弧相当于沿纤维轴方向周期约为0.34nm 的重复结构所产生的衍射，揭示了 DNA 分子中含有大量的以某种型式存在的规则结构。1953年，R. E. Franklin[77]拍摄到十分好的照片，成为 Watson-Crick 的右手双螺旋 DNA 模型的主要实验根据。她将足量的蒸馏水加至小块纤维状的小牛胸腺 NaDNA 固体中形成胶，将一针尖放入胶中，然后慢慢地抽拉。由适当地控制胶的含水量和抽拉的速度，可以随意将 NaDNA 胶拉成直径为小到不足1μm 直至大到100μm 的纤维。将直径为50μm 的单根纤维用纤维素胶直接固定在直径为100μm 的 X 射线准直器的管口上，拍摄纤维衍射照片，用 Ni 滤波的 CuK_α 射线，样品与底片的距离为15mm，照射62小时。由于这样小的样品的衍射十分弱，为了消除由空气产生的散射，在整个实验中，将氢气通过合适的无机盐饱和溶液后进入 X 射线衍射照相机内，并在相机内放置一个盛有相同饱和盐溶液的小容器。样品在照射前需在这个相机中放置至少1小时。用这样的方法，使整个照相过程在恒定的湿度下进行，以控制



(a) C3′-内型糖折叠　　(b) C2′-内型糖折叠

图8.5　核酸中碱基堆积时的"顺式"构象和"反式"构象

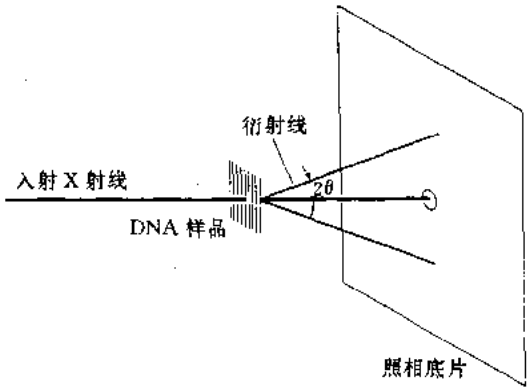


图8.6　DNA 纤维样品的衍射原理图

样品的含水量。结果得到了十分好的衍射照片，见图8.7(b)。DNA 的纤维因含水量和阳离子不同而分别产生 A 型 DNA[图8.7(a)]、B 型 DNA[图8.7(b)]以及 C 型 DNA(图8.7c)三类衍射花样。A-DNA、B-DNA 以及 C-DNA 分别是相对湿度为75%的 NaDNA、相对湿度为92%的 NaDNA 以及相对湿度为66%的 Li DNA。它们的衍射花样，特别是 B 型，在子午线区域中呈现出螺旋结构所特有的衍射花样，即衍射斑点的 X 形排列，并且在 B 型衍射花样中有一个相当于0.34nm 周期结构的十分强的子午线衍射斑点。还发现 A 型与 B 型可相互转化。

M. H. F. Wilkins[78,79]也进行了许多研究工作，除拍摄了上述 C 型 DNA 即 Li DNA 的衍射照片并对 A-、B-和 C-DNA 模型进行了结构修正外，还对从大肠杆菌、小麦胚芽以及猪胸腺等提取的 DNA 进行了衍射实验，发现都给出 B 型衍射花样。

8.2.2　DNA 右手双螺旋结构模型

众所周知的 Watson-Crick 右手双螺旋 DNA 模型是在上述 DNA 的 X 射线纤维衍射研究的基础上建立起来的。前人进行了大量的工作，积累了丰富的资料，并曾提出了多种 DNA 结构模型。其中，R. E. Franklin 曾指出 DNA 是两条链的螺旋，不过她没有给出具体模型。当时也

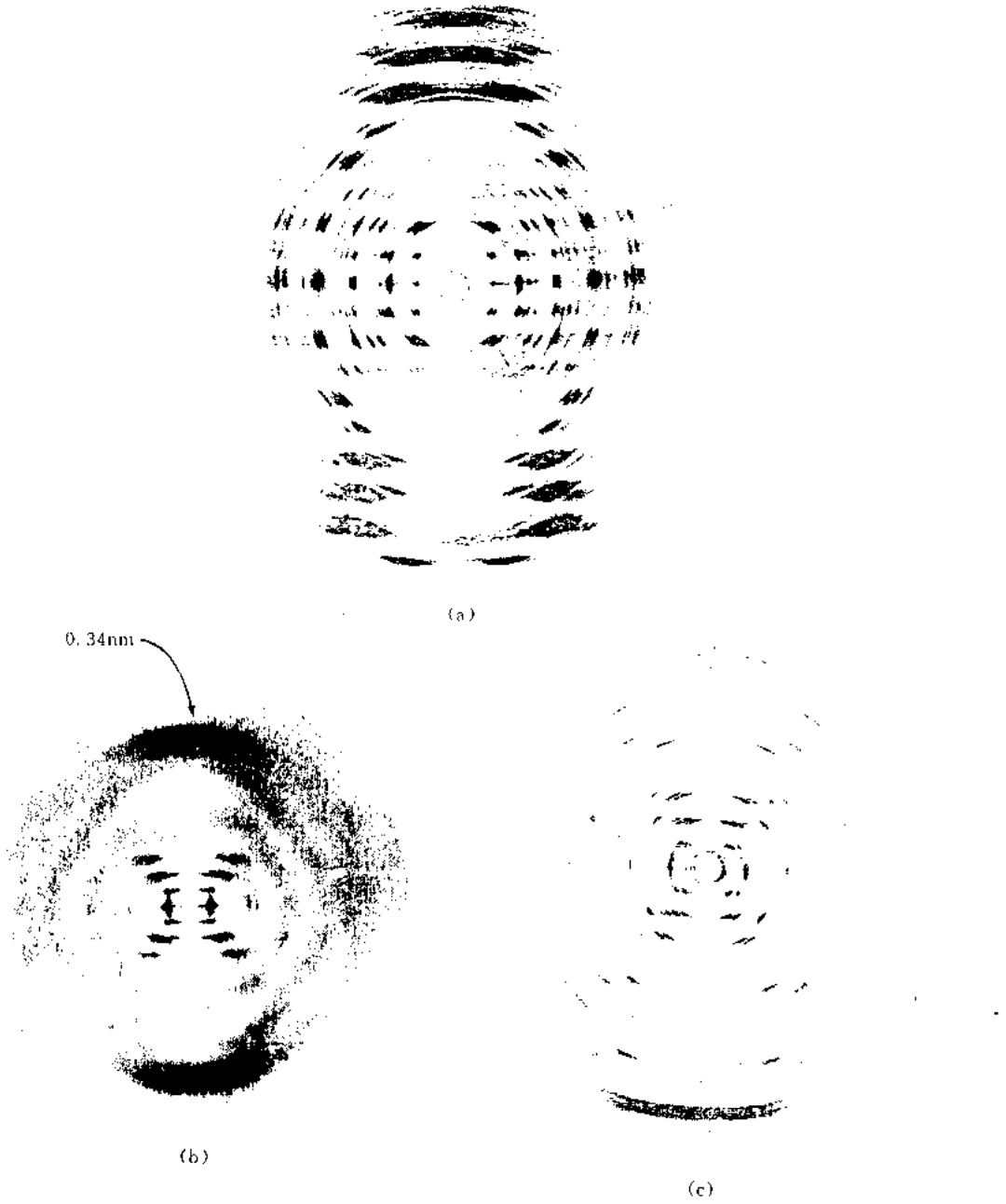


图8.7　A 型(a)、B 型(b)和 C 型(c)DNA 的纤维衍射花样

有人发表了另一些具体的螺旋结构模型：在一个模型中，DNA 是由三股相互缠绕的链组成的，磷酰基靠近纤维轴，碱基在外面；在另一个三股链结构模型中，磷酰基在外面，碱基在内部由氢键结合在一起；还有四股链模型等。J. D. Watson 和 F. H. C. Crick 否定了上述模型，指出第一个三股链模型至少有两点不合理，其一是由于产生 X 射线纤维衍射花样的物质是 DNA 盐，而不是 DNA 自由酸，所以没有酸性的氢离子，无法解释什么力量能使结构维系在一起，特别是带负电荷的磷酸盐若靠近轴，会相互排斥；其二是有一些范德华距离太小了。对于另一个三股

链模型和四股链模型等，则认为显然难以解释，特别是对于当时生物化学家已经阐明了的关于碱基A与T以及G与C含量分别相等的这一基本事实都被忽略了。Watson和Crick为DNA盐提出了一个根本不同的模型，于1953年发表在"自然"杂志上，这就是著名的Watson-Crick右手双螺旋模型[75]，见图8.8。

图8.8 Watson-Crick右手双螺旋DNA模型
这是1953年发表在自然杂志上的模型。两条带表示两条磷酸盐-糖链，水平方向的棒表示将两条链支撑在一起的碱基对

Watson和Crick依据B-DNA纤维衍射花样中相当于0.34nm周期结构的子午线反射以及碱基A与T，G与C含量相等的定量分析，根据通常的立体化学原理和数据，用搭模型的方法，正确地预言出DNA的立体结构，并解释了遗传机理。在模型研究中，他们注意到在A与T和G与C之间能形成以氢键相连接的碱基对，图8.9为后来经M. H. F. Wilkins[79]精修过的碱基对结构图，两种碱基对具有相同的尺度。DNA作为一个规则的两条链的结构，在两条链上相对着的碱基之间形成氢键。细致的模型建造研究表明了，只有当两条链是反平行时，才能形成这样的碱基对结构，见图8.10。然而，反平行排列的两条链具有形成各种类型构象的可能性。一种可能的构象是梯子形，见图8.11(a)。在其中，各链将是完全伸展的，因而残基之间在长轴方向上的距离应为0.68nm。相邻碱基对的距离也是0.68nm，这样，就会在它们之间形成大的空隙，必须由水分子来填补。但是，这将是能量上不利的结构，因为碱基平面是疏水性的，它们倾向于与其他相似的疏水表面接触。反平行双链的另一种可能的构象是由梯形结构以简单的右手扭曲而形成的双螺旋结构。当它形成时，碱基对之间的距离减小，直至它们以0.34nm的距离而堆积在一起，使相邻碱基对之间形成最紧密的接触。在这种构象中，大约每10个核苷酸残基即沿螺旋轴每3.4nm，结构重复出现，这能很好地解释衍射花样中3.4nm和0.34nm的衍射斑点。图8.11(b)为这种构象的示意图。

在Watson-Crick的DNA模型中，B-DNA分子是两条反平行的多核苷酸的右手双螺旋结构。两条链以垂直于螺旋轴的二重轴对称性缠绕在一起，带负电荷的磷酸-糖主链在分子的外部，阳离子容易接近它们，磷原子与螺旋轴相距1.0nm，因而分子的厚度约为2.0nm；碱基对(A—T，G—C)在分子的内部，以氢键相连系，所有碱基对以0.34nm的间距堆积在与螺旋轴垂直的平面上；糖环大致垂直于与其相连接的碱基平面；每10个碱基对形成一个完全的周期，同一条链上相邻核苷酸残基之间相差36°，因而每10个核苷酸残基之后结构重复，即分子在螺旋轴方向上的周期为3.4nm。

如果用核糖代替脱氧核糖，那么就不能建造成这个模型，因为多出的那个氧原子会使范德华距离太近了；另外，还发现假若碱基在结构中仅以最常见的酮式(而不是烯醇式)存在的话，那么只有A与T或G与C才能结合在一起。因此，虽然在单链上碱基的序列是不受任何限制的，但是只要一条链上的碱基序列已确定了的话，另一条链上的碱基序列自然地也确定了。Watson和Crick从这种特殊的碱基配对结构指出了遗传学物质可能是通过DNA拷贝的机制进行遗传信息传递的。他们坚信DNA就是基因，这在当时很多人还是持怀疑态度的，这个模

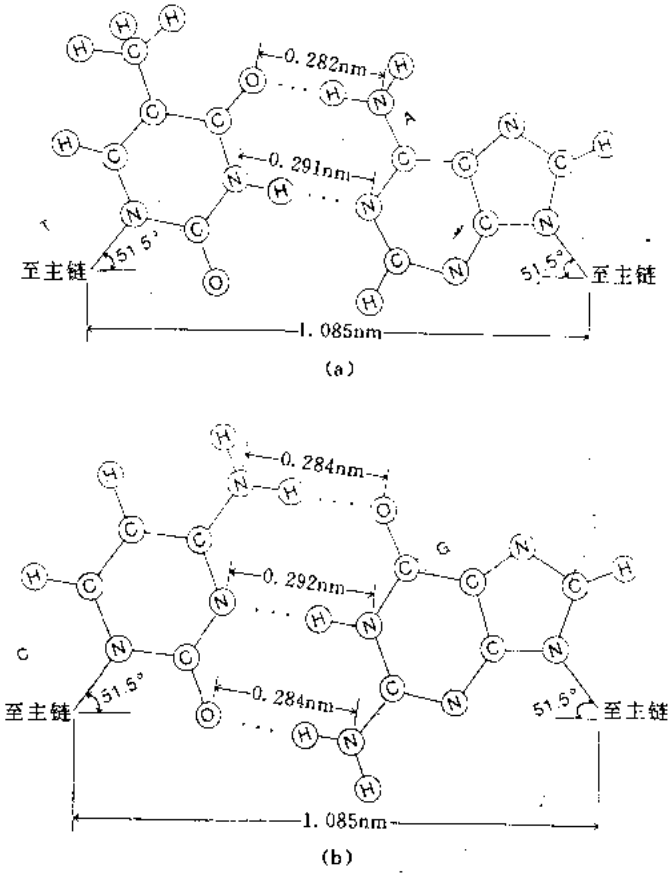


图8.9　两种碱基对的氢键和尺度

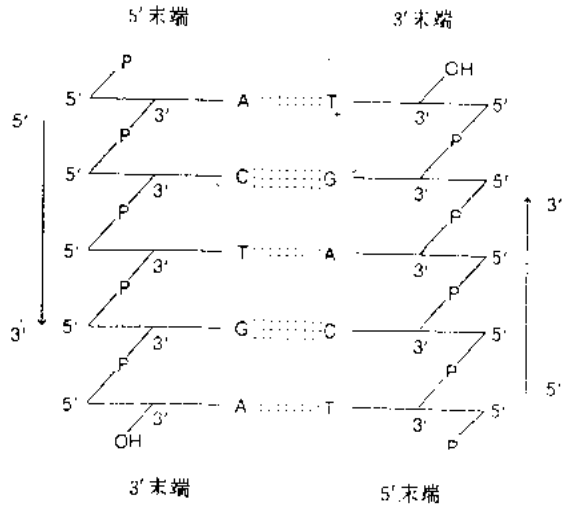


图8.10　碱基互补的反向双股 DNA 片段示意图

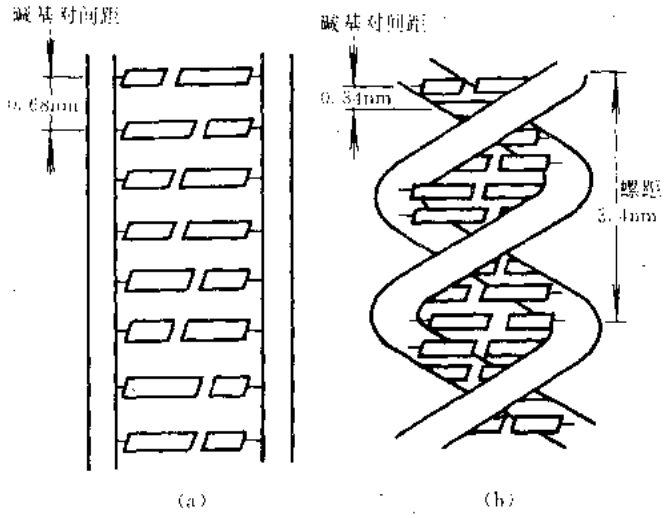


图8.11 碱基配对的DNA的两种可能的不同构象的示意图
(a)梯形构象;(b)双螺旋形构象

型提出后,基因及复制等问题迎刃而解。这个模型提出9年以后,他们被授予1962年诺贝尔生理学和医学奖。几十年来,遗传密码、DNA重组、DNA调控、遗传工程及蛋白质工程等都迅速被阐明、发展或应用了。

根据DNA纤维衍射花样,用与B-DNA相似的方法,人们推测出其他类型DNA的右手双螺旋模型。在A-DNA模型中,分子的周期是2.82nm,每个周期中有11个碱基对,这些碱基对的平面以大约19°倾斜于分子的轴;在C-DNA模型中,分子的周期为2.88nm,每个周期中有9.3个碱基对,它们的平面以6°倾斜于分子的轴;此外,还发现了其他类型的DNA构象。例如,在T型DNA[80]中(从T_2噬菌体得到的DNA),分子的周期为2.72nm,一个周期中有8个碱基对。

8.2.3 人工合成寡核苷酸的结晶和单晶X射线衍射结构分析

DNA是遗传物质,因而了解它的细微结构变得越来越重要,尤其是酶对DNA的作用以及基因调控中涉及到许多蛋白质,都要回答一个关键问题,即蛋白质所识别的DNA的局部区域是否具有与DNA其他部分不同的结构,以解释DNA-蛋白质相互作用的特异性机理等等。

由于纤维衍射所产生的衍射点数目太少,实验数据不足,因而只是晶体学分析的一个粗略的方法。结构分析受到限制,只能提供DNA分子的平均构象,不能分析至原子水平,因而不能测定由特定的碱基序列所引起的在DNA双螺旋中的局部变化。与蛋白质的精确结构测定相同,要求具有三维有序性的单晶体,只有X射线单晶衍射法才能获得大量衍射数据,精确测定出DNA的结构。但是,DNA的分子量太大,无法生长单晶。有人试图用天然DNA的片段生长晶体,至今也未见成功的报道。直到20世纪70年代末,人工合成DNA片段的技术已发展到能有效地制备X射线单晶衍射法所要求的量和纯度时,DNA的高分辨率结构分析才得以实现。

用人工合成的自身互补寡核苷酸片段能较好地模拟天然DNA双螺旋结构。1978年[81]发表了第一个被结晶和进行了单晶结构分析的这种寡核苷酸片段,即四核苷酸d(pApTpApT)。由于脱氧寡聚核苷酸的单晶X射线衍射分析的主要目的是试图揭示碱基序列对DNA双螺旋

结构的影响，并由此而了解与蛋白质识别 DNA 有关的碱基对所引起的构象变化，因此，人们所设计合成的 DNA 片段常为含有已知的与蛋白质相互作用的特殊的碱基序列。例如，d(CGCGAATTCGCG)所含有的 GAATTC 序列，是 EcoR I 限制性内切酶的天然识别和切割的位点，而同时测定 d(CGCGAATTCGCG)的溴化物的晶体结构是为了模仿在基因表达调控过程中所涉及到的5-甲基胞嘧啶所引起的结构变化。又例如，d(GGTATACC)含有的 TATA 序列，是原核生物和真核生物基因的启动基因区域中的一个交感序列部分，并与在 DNA 双螺旋解缠绕和进行转录之前的 RNA 聚合酶的结合有关。再如，d(GGCCGGCC)含有的 CCGG 是限制性内切酶 Hpa Ⅱ 的识别位点等等。

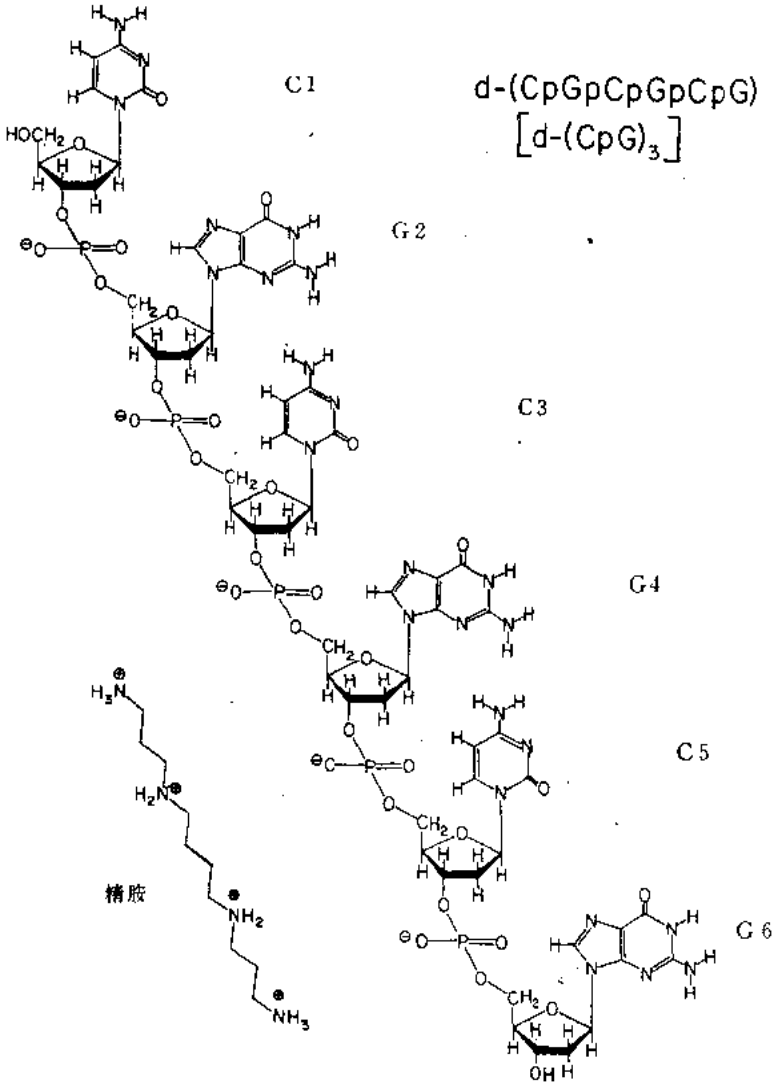


图8.12　多聚胺类化合物在 DNA 和 RNA 结晶过程中的反离子作用

与蛋白质类似，DNA 片段作为生物多聚分子的结晶作用与浓度、溶剂的种类、溶液的离子强度和 pH 以及温度等都有关系。而且，根据 t-RNA 晶体生长的经验(见8.3.1)，在生长 DNA 片段的晶体时，也常需加入一些多聚胺，例如精胺、尸胺、精脒或腐胺等，它们可能是作为磷酸

根负离子的反离子而起作用的，见图8.12。现以1979年由 A. Rich 的实验室完成的著名的左手

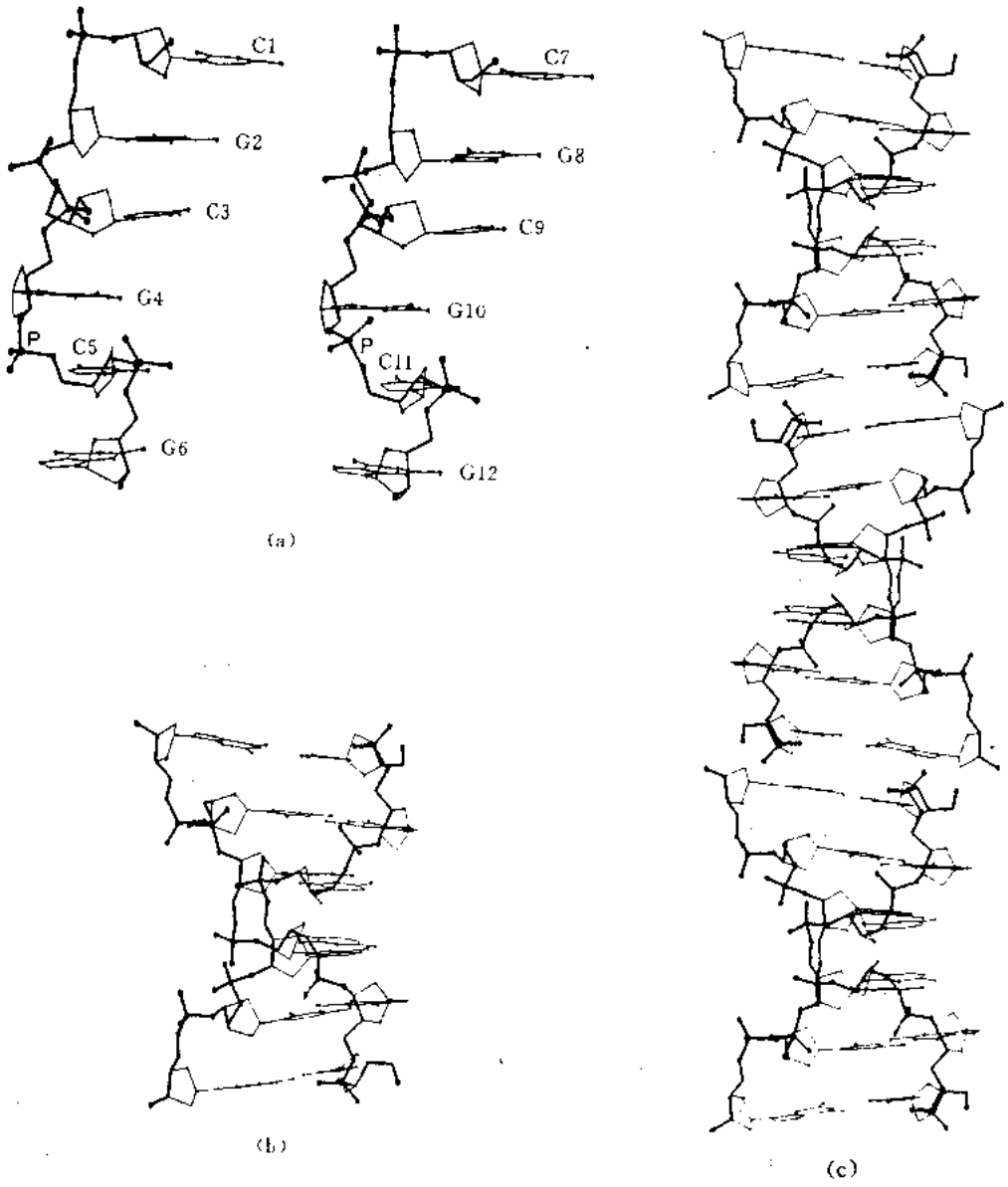


图8.13 Z-DNA 的结构

(a)两个独立的六核苷酸 $d(CG)_3$ 的构象；(b)两条 $d(CG)_3$ 链反平行互补形成一个双螺旋分子；
(c)三个双螺旋分子沿晶胞的 C 轴堆积，看起来像一个连续的双螺旋结构

双螺旋 DNA 片段 d(CGCGCGp)的结晶作用为例：用坐滴式蒸汽扩散法，以5%异丙醇/水作沉淀剂，用 $d(CG)_3$ 的铵盐在含有30mmol/l 二甲胂酸钠盐缓冲液(pH7.0)、10mmol/l 精胺四氯化物、15mmol/l$MgCl_2$ 以及2mmol/ld $(CG)_3$ 的溶液中生长晶体。4℃下3个星期，得到了0.7mm×0.7mm×0.5mm 的晶体。像蛋白质晶体那样，在进行 X 射线衍射实验时，需将 d $(CG)_3$ 的晶体装入封闭的玻璃毛细管中，并放入一滴母液。晶体衍射分辨率达到了0.09nm，共收集到15000以上个衍射点的强度数据。制备了三个不同的重原子衍生物(Ba^{2+}，Co^{2+} 以及

Cu^{2+})，用多对同晶置换法解出了原子水平的结构(见图8.13)。

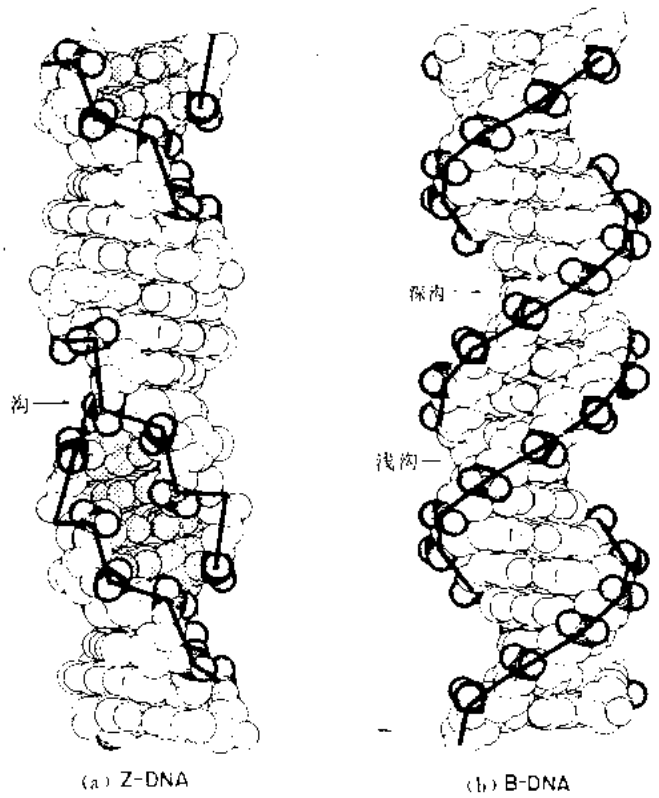


图8.14 Z-DNA(a)与B-DNA(b)结构外观的比较

现在，单晶衍射分析已成为研究DNA立体结构的主要手段。有的已测至原子水平，有的虽受到晶体衍射能力的限制，分辨率仅达0.25nm至0.18nm，但是主要的构象特点如碱基平面所取的扭角等可以被测定到几度的精确度，甚至各个原子的位置也能被识别。至今已公布了几十种人工合成DNA片段的晶体结构。除了大多数均为右手双螺旋DNA构象外，如前所述，1979年A. Rich等人[82]还首次发现d(CGCGCG)为左手双螺旋DNA。此外，1980年由R. E. Dickerson的实验室[83]对自身互补的d(CGCGAATTCGCG)12聚体单晶进行的高分辨率X射线衍射结构分析，也首次确证了通常被认为在生物体中占优势的B-DNA构象即Watson-Crick模型的基本正确性。特别是，现在所观察到的所有DNA构象，包括A-DNA，B-DNA，C-DNA以及Z-DNA等，都保留了Watson-Crick模型的碱基对氢键类型。但是，至今研究过的大多数DNA片段具有A-DNA构象。这表明A-DNA构象在相当广泛的条件下是稳定的。很可能在生物体中，A型和B型都是重要的。此外，从现有的研究结果，已可得出下述结论：(1)在双螺旋的构象中，有局部的可变性；(2)碱基堆积的形式是这些局部构象变化的主要决定因素。由于糖-磷酸主链和糖苷键的扭角变化使碱基以能量有利的各种形式堆积；(3)局部的变化能导致双螺旋结构整体的变化如弯曲和沟的加宽等。局部的和整体的构象变化都可能为DNA被某种蛋白质识别提供特异性。由于目前所测定的DNA片段数目有限，因此还没有足够的资料断定DNA构象的变化是决定于特殊的碱基序列而不是外部条件。但是可以预料，当更多的

DNA 片段、DNA 结合蛋白和它们的复合物以及调控蛋白的晶体结构在原子水平上被分析以后，所发现的 DNA 结构的多样性将会对生物学的发展具有深远的意义。

8.2.4 Z-DNA 左手双螺旋结构

前面已经提到，1979年 A. Rich 等人用人工合成的寡核苷酸 d(CpGpCpGpCpG)进行单晶衍射结构分析，发现了一种左手双螺旋 DNA 构象，见图8.13。图中可以看到两条反向的 $d(CG)_3$核苷酸链互补形成一个左手双螺旋分子，三个这样的分子沿着晶胞的 C 轴堆积，虽然在这个螺旋结构中由于在每个12体分子的末端残基上缺少一个磷酸根基团而被中断，但看起来像一个连续的左手双螺旋结构。在这种螺旋结构中，由于糖苷键的扭角是顺反交替的(在 B-DNA 中全为反式的)，使得糖-磷酸主链的走向成为 Z 字形，即连接磷酸基团的线为 Z 字型(见图8.14)，因而被称为 Z-DNA(在 B-DNA 中，连接磷酸基团的线是光滑的)。在 Z-DNA 中，螺旋的重复周期即螺距为4.5nm，含有12个碱基对，即 $d(CG)_3$的两个双螺旋分子；因而每个核苷酸残基上升距离为0.37nm；碱基对的平面以7°倾斜于分子的轴。

现在已有证据表明，Z-DNA 也可存在于天然 DNA 的特殊区域中，这种不常见的构象可能在调控基因表达中起着重要的作用。

8.2.5 DNA 双螺旋结构的构象变化

综前所述，在所有天然存在的和人工合成的 DNA 双螺旋结构中，都发现了相同的碱基对堆积结构。但是，由于每个核苷酸具有6个可旋转的单键以及糖环具有一定的柔韧性，可以形成4种折叠构象，因而实际观察到了前述各种不同构象的 DNA 双螺旋结构[84]，在表8.2和图8.15中进行了比较。

表8.2 A-，B-和 Z-DNA 结构的比较

	A-DNA	B-DNA	Z-DNA
螺旋	右手的	右手的	左手的
每圈的残基数	11	10	12(6个二体)
每个残基上升的距离(nm)	0.255	0.34	0.37
螺距(nm)	2.8	3.4	4.5
碱基对倾斜度(°)	20	6	7
每个残基旋转的角度(°)	33	36	−60(每个二体)
糖苷的构象			
脱氧胞苷	反式	反式	反式
脱氧鸟苷	反式	反式	顺式
糖的折叠			
脱氧胞苷	C-3′-内型	C-2′-内型	C-2′-内型
脱氧鸟苷	C-3′-内型	C-2′-内型	C-3′-内型

在 B-DNA 中，螺旋每圈含10个碱基对，碱基对的平面接近于与螺旋轴垂直，糖环为 C-2′-内型构象。B-DNA 被认为是 DNA 在溶液中的主要构象。当水化的 DNA 纤维中失去一些水后，双螺旋压缩靠近，变成了 A-DNA 构象，每圈大约为11个碱基对，碱基对倾斜于螺旋轴约20°，糖环的折叠变为 C-3′-内型构象。Z-DNA 是左手双螺旋，每圈含12个碱基对，碱基对的平面相对于 B-DNA 中的碱基对平面为绕糖苷键旋转了大约180°，见图8.16。其中鸟嘌呤绕糖苷键的旋转使脱氧鸟苷变为顺式构象；而对于脱氧胞苷来说，由于胞嘧啶和脱氧核糖都被旋转

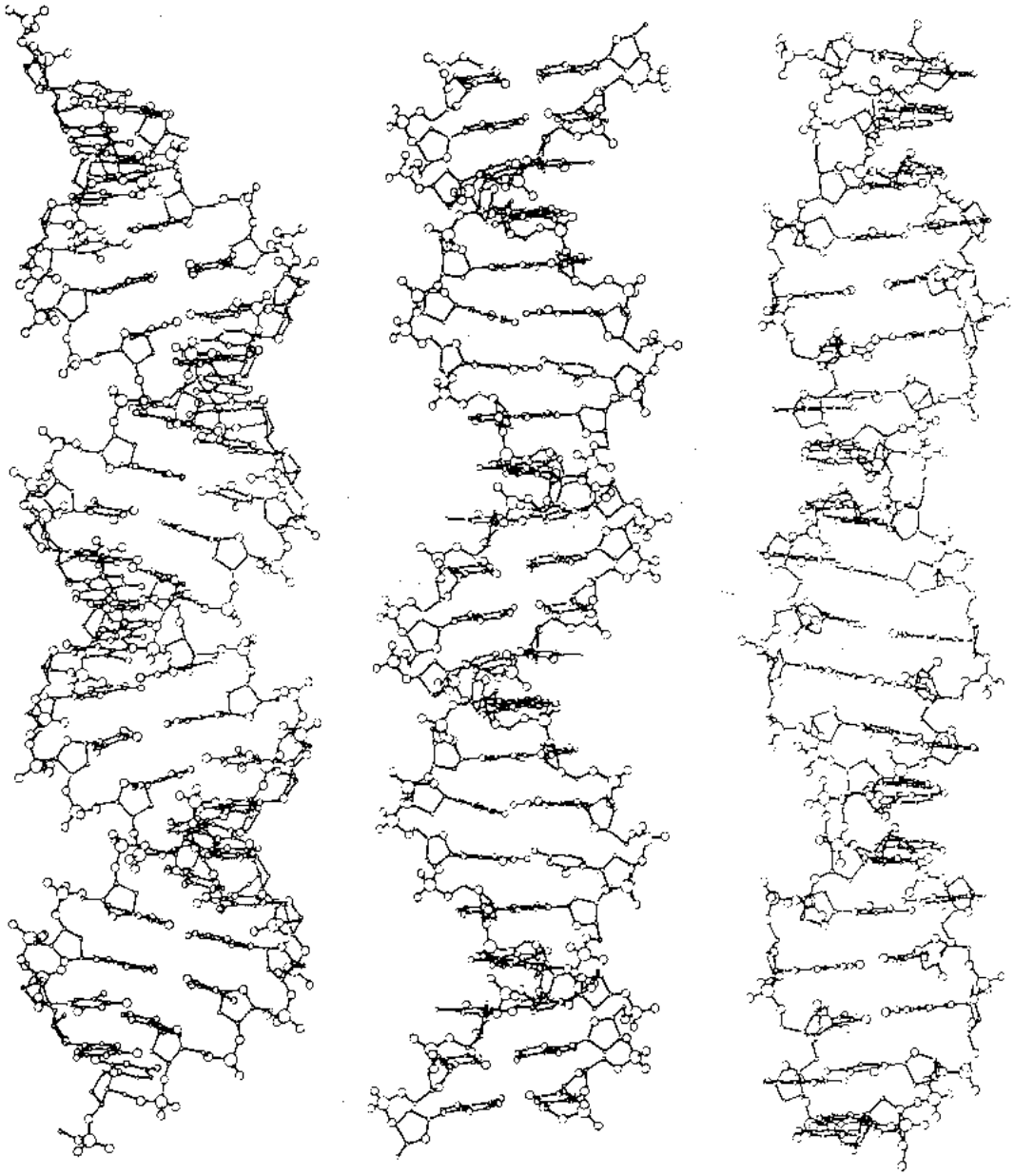
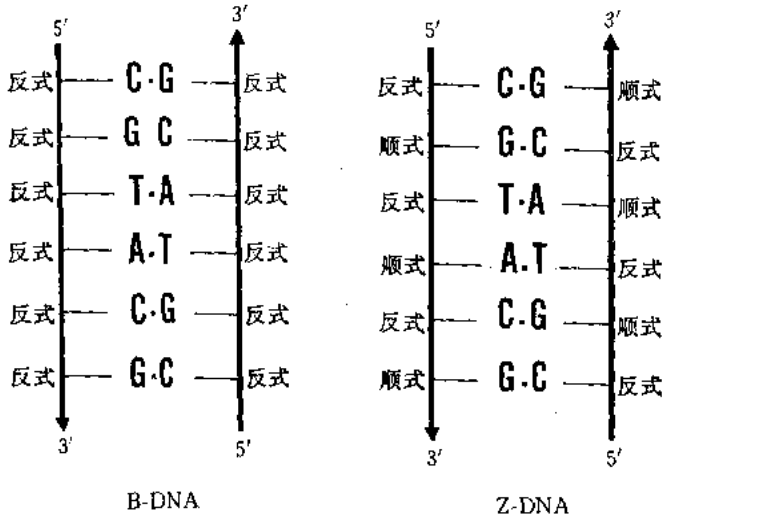


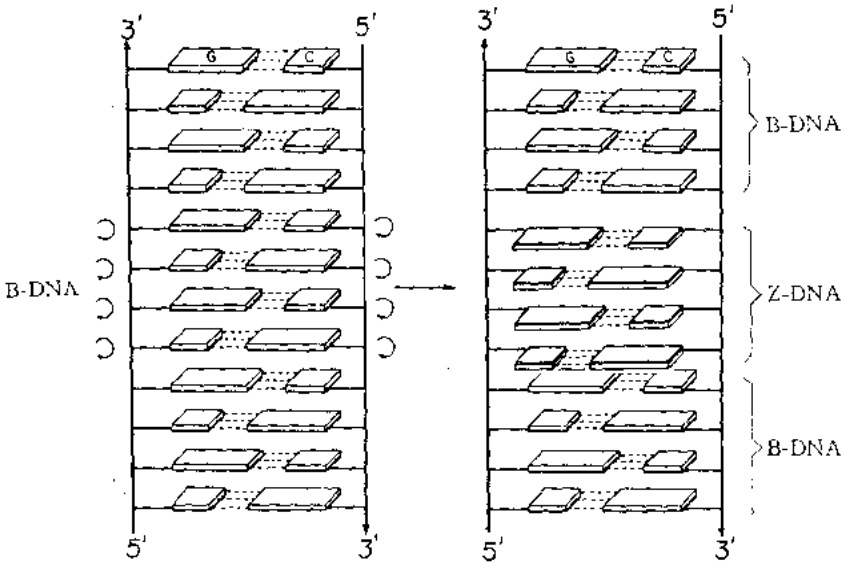
图8.15 由人工合成寡核苷酸的晶体结构所模拟的无限伸展的 DNA 结构

(a)由八核苷酸(GGTATACC)的中央6个碱基对进行重复而产生的序列为…(GTATAC)$_n$…的 A-DNA;(b)由十二核苷酸(CGCGAATT^{Br}CGCG)的中央10个碱基对重复而产生的序列为…(GCGAATTCGC)$_n$…的 B-DNA;(c)由六核苷酸(CGCGCG)的中央4个碱基对重复而产生的序列为…(GCGC)$_n$…的 Z-DNA

了,因而仍保持反式构象。模型建造表明,顺式构象比反式构象在空间中较为拥挤;嘌呤核苷酸比嘧啶核苷酸较易形成顺式构象。在 Z-DNA 中脱氧鸟苷和脱氧胞苷所采取的顺反不同构象,使得糖-磷酸盐骨架成为 Z 字形。而且,正因为嘧啶难以形成顺式构象,所以很可能只有在一级结构具有交替的嘌呤和嘧啶序列时才能形成 Z-DNA。不过,现在的观察也发现左手螺旋的



(a)



(b)

图8.16 B-DNA 和 Z-DNA 的拓扑学关系变化示意图

假若将 B-DNA 中的4个碱基对绕糖苷键旋转约180°,则变成了 Z-DNA 片段,这些碱基对将不能直接堆积在 B-DNA 片段中的碱基对上

DNA 也可以存在于天然 DNA 的特殊区域中。

8.3 核糖核酸(RNA)的晶体结构

现已发现在生物细胞中含有三种主要的 RNA,即核糖体 RNA(rRNA)、转移 RNA(tRNA)以及信使 RNA(mRNA)。其中 rRNA 由～100--3100个核苷酸组成,约占全部 RNA 的80%,是构成细胞中执行蛋白质合成的颗粒即核糖体(ribosomes)的骨架结构;tRNA 由

75—90个核苷酸组成，占16%，它能将激活了的各种氨基酸携带至mRNA模板上，每种氨基酸都有一种或几种与其相应的tRNA；mRNA由75—3000个核苷酸组成，仅占2%，其主要功能是将贮存于DNA中的遗传信息转运到核糖体中。此外，还有病毒RNA等。

RNA与DNA的一级结构及其异同已在前面讨论过了。近年来，试图研究各种RNA晶体结构的人越来越多了，但是由于RNA比蛋白质更难以得到晶体，所以尽管已有数百个各种RNA的核苷酸序列已被测定，但空间结构却很少被测定。到目前为止，用X射线衍射法研究得较清楚的只有t-RNA的晶体结构。关于mRNA，至今尚无二级结构和三级结构的可靠资料，这是由于它们很不稳定，而且有证据表明它是与特殊的蛋白质结合在一起而存在的，因而要将mRNA进行结晶和用X射线衍射法测定结构是困难的。至于rRNA的空间结构测定，已有一些进展，例如高嗜热细菌HB8的5SrRNA已生长出若干种晶体并进行了低分辨率X射线衍射研究。此外，还有病毒中的RNA将在第十章中讨论。本章只讨论tRNA的晶体结构。

所有的tRNA，不论来自动物、植物还是微生物，分子量都在25000道尔顿左右，由70—90个核糖核苷酸残基组成。其5′-末端大多为pG…，也有为pC…的；其3′-末端均为…CpCpAOH。t-RNA的二级结构都呈三叶草形。由于双螺旋结构所占的比例较高，所以是十分稳定的。三叶草结构由氨基酸臂(Amino acid arm)、二氢尿嘧啶环(DHU-loop)、反密码子环(Anticodon loop)、额外环(Extra loop)和TΨC环(TΨC loop)等五部分组成，见图8.17。从图中可以看到，t-RNA分子中含有十多个稀有碱基。

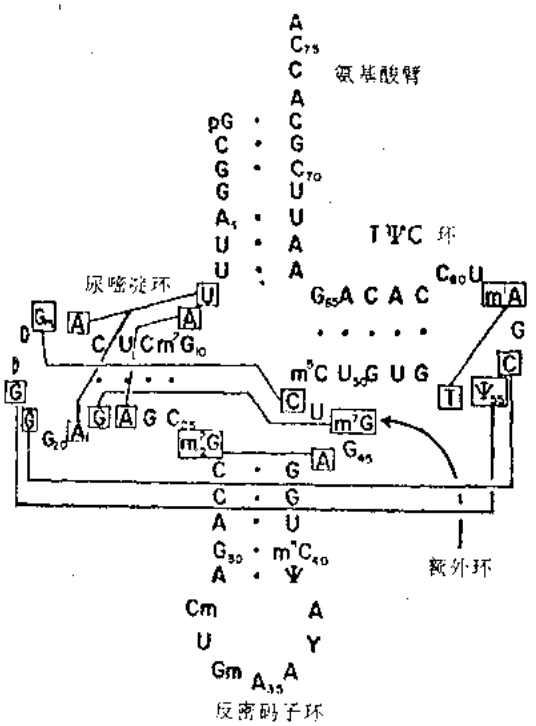


图8.17 酵母 $tRNA^{Phe}$ 的核苷酸序列(三叶草形结构)及二级结构的相互作用
连接碱基的·和实线表示氢键

8.3.1 RNA的结晶和晶体结构分析

由于tRNA分子量较小，因此，可以生长出晶体，并用X射线单晶衍射法进行高分辨率的结构分析。现已报道生长出晶体的有从酵母得到的t-RNA混合物、t-RNA^{Phe}，t-RNA^{Ser}，t-RNA^{Val}，t-RNA^{Gly}和t-RNA^{Asp}等以及从大肠杆菌得到的t-RNA，t-RNA^{Gln}，t-RNA^{Leu}、t-RNA^{Fmet}，t-RNA^{Phe}和t-RNA^{Tyr}等。这些t-RNA的晶体，大多数是用汽相扩散法获得的，而且广泛应用了挥发性有机溶剂，如乙醇、二氧六环、异丙醇、丙酮以及2-甲基-2，4二羟基戊二醇(MPD)等作沉淀剂。由于t-RNA对小离子的敏感性可能比大多数蛋白质大得多(这也许是反映了它的外表面有特殊的阴离子的特性)，因此在生长晶体时，常需加入一些小的多聚胺如精胺、尸胺、精脒和腐胺等，并被认为是生长高分辨率X射线衍射分析用的优质晶体的关键性成分。这些多聚胺可能是对磷酸根负离子起着特异的反离子作用。另外，金属阳离子如Mg^{2+}，Ca^{2+}，Co^{2+}以及Mn^{2+}等常为RNA结晶作用所必需的。而且金属阳离子与多聚胺的浓度之比是一个决定性的因素。这一原理后来也应用于人工合成的DNA片段的晶体生长(见8.2.3)。以第一个测定出晶体结构的酵母t-RNA^{Phe}正交晶体生长条件为例[85]：用坐滴式汽相扩散法，以10%异丙醇或10—12%2-甲基-2，4-二羟基戊二醇(MPD)为平衡溶液；结晶溶液中含有3—5mg/ml t-RNA，10mmol/l二甲胂酸钠盐缓冲液(pH6.0)，10mmol/l $MgCl_2$以及1mmol/l精胺盐酸盐。在4℃下生长晶体，通常需几天至长达2个星期。晶体和旋进照片见图8.18。空间群

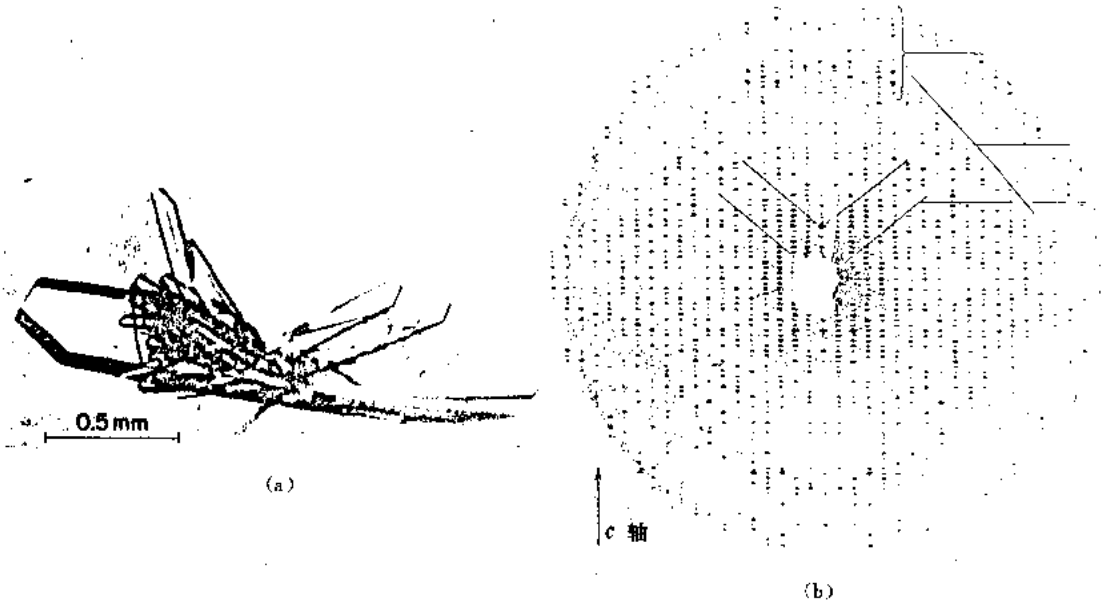
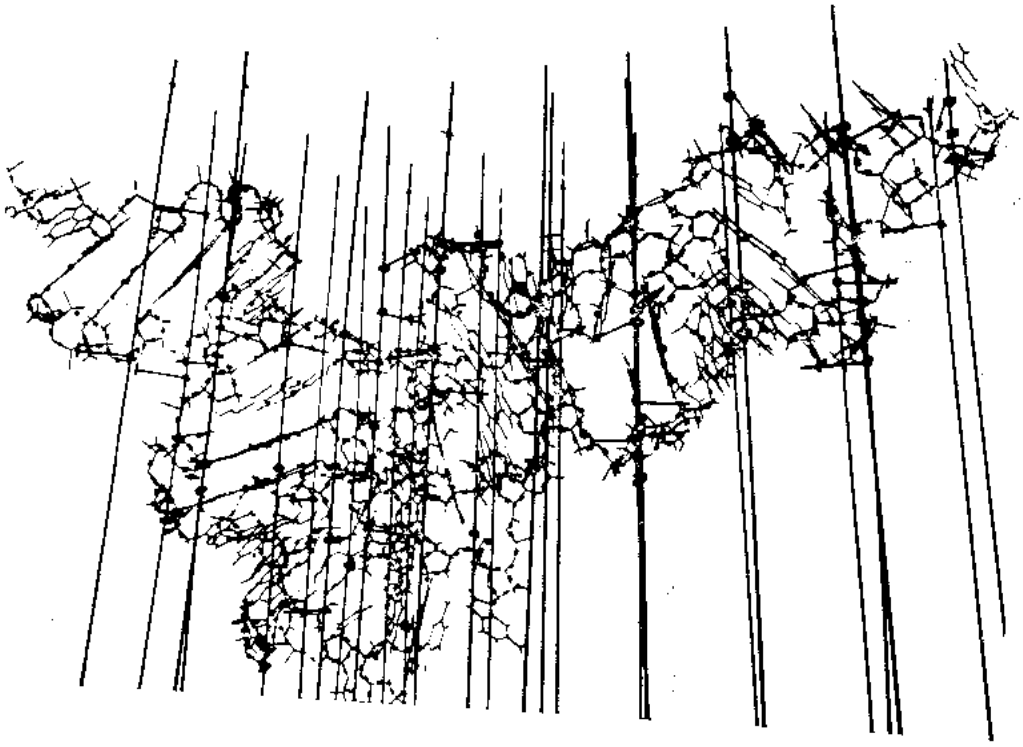


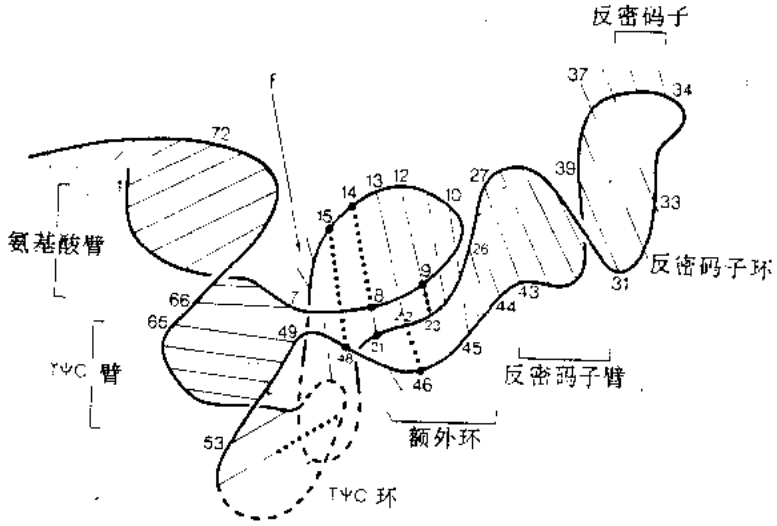
图8.18 酵母tRNA^{Phe}的正交晶体(a)和旋进照片(b)

为$P2_122_1$，晶胞参数为：$a=3.3nm$，$b=5.6nm$，$c=15.1nm$。制备了重原子衍生物，如ATP-Os，Pt，Sm等，用多对同晶置换法分析晶体结构[86]。

除了天然t-RNA晶体结构研究外，80年代还开始了对人工合成的寡聚核糖核苷酸进行单晶X射线结构分析，例如(UCAUAUAUAUAUAA)等。



(a)



(b)

图8.19 酵母 t-RNA^{Phe}的晶体结构(a)和示意图(b)

8.3.2 t-RNA 的晶体结构

1974年，A. Rich[86]小组和 A. Klug[87]小组分别报道了酵母 t-RNA^{Phe}正交晶体和单斜晶体 0.3nm 分辨率的结构，两者基本相同，但在维持三级结构的某些相互作用力方面差别较大。

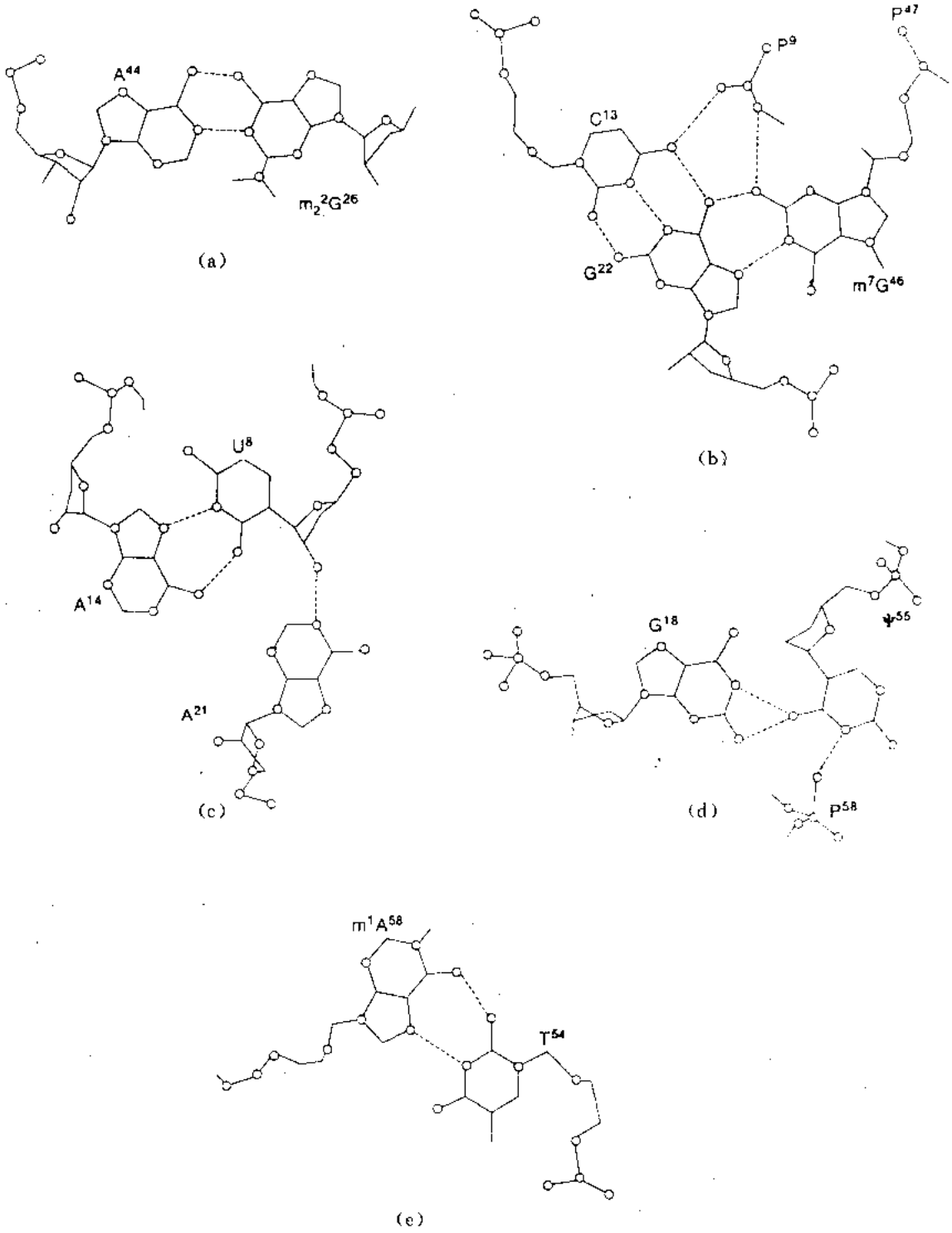


图8.20　酵母 t-RNA^{Phe}中的一些不常见的使三级结构稳定的氢键的例子

图8.19是 A. Klug 小组测出的结构。t-RNA^{Phe}分子呈 L 形，分子线度为6.5×7.5×2.5nm。氨基酸臂和 TΨC 臂形成一个连续的双螺旋区，构成字母 L 的下面一横，而二氢尿嘧啶臂与其垂直，它与反密码子臂及反密码子环共同构成字母 L 的一竖。反密码子臂经额外环而与二氢尿嘧啶臂相连接。此外，二氢尿嘧啶环中的某些碱基与 TΨC 环及额外环中的某些碱基之间形成一些额外的碱基对。这些额外的碱基对是维持 t-RNA 三级结构的重要因素。

现有资料表明，Watson-Crick 碱基对类型在所有天然的核酸中包括 RNA 在内都是普遍存在的。在 tRNA^{Phe}分子中，有20个碱基对是以 Watson-Crick 型式的氢键相连的；还有另外40个左右的氢键不属于这种类型，其中有些是不常见的、在 t-RNA 的不同区域之间形成的氢键，它们使复杂的三级结构变得稳定。图8.20中可以看到一些以这种不常见的氢键所连系的三级结构的例子。除此以外，在 t-RNA^{Phe}中，磷酸基团和核糖的 C-2′-OH 基团也参与了氢键。但是，在 DNA 中已经讨论过的关于碱基堆积对于使规则的核酸双螺旋稳定的重要性，在t-RNA中也同样是重要的特点。在酵母 t-RNA^{Phe}分子的76个碱基中，只有4个(D^{16}，D^{17}，U^{47}和 G^{20})未参与堆积，其中 D^{16}和 D^{17}二氢尿嘧啶残基在任何条件下都不可能堆积，因为它们不再具有其他碱基所具有的平面的构象。可以认为，在 t-RNA 中，碱基的堆积和各种氢键对于稳定它的构象具有大致相等的贡献。

与 DNA 中的脱氧核糖不同，在 RNA 中为核糖，即在 C-2′位置上有一个—OH 基团，它与 C-3′—OH 基团形成顺式构型。很可能由于 DNA 中的糖环上无 C-2′—OH 而使得 DNA 较之 RNA 来说具有较大的化学稳定性。在 RNA 中，核糖倾向于形成 C-3′-内型折叠构象，因而 RNA 中的双螺旋具有类似于 A-DNA 的构象，甚至于在高度水化的情况下也如此。

从 t-RNA^{Phe}的晶体结构中可以清楚地看到它的三维空间结构与其生物学功能的密切关系。在分子的反密码子环上的3个反密码子碱基，例如 tRNA^{Phe}的 AAG(3′→5′方向)朝外翻出，它可互补地结合到核糖体中的 mRNA 的密码子例如 UUC(5′→3′方向)上；而分子另一端的氨基酸臂则与一个特定的氨基酸结合，例如 t-RNA^{Phe}可与 Phe 结合，从而将这个氨基酸携带至 mRNA 模板处，进行按特定的氨基酸序列合成多肽链的过程。

第九章　病毒的立体结构

病毒是由蛋白质和核酸组成的最简单的、非细胞形态的生物。有些病毒还含有脂类、糖类物质以及金属离子等。在病毒颗粒中，核酸位于内部，蛋白质在外部包裹着核酸，形成蛋白质外壳。每个病毒颗粒含有一种或一种以上的蛋白质，每种蛋白质通常由许多相同的亚基组成。病毒含有 RNA 或 DNA，两者不可兼有。有的病毒只含有 RNA（单链的或双链的），称为 RNA 病毒；有的只含有 DNA（所有 DNA 都是双链的，其中有些是环状的 DNA），称为 DNA 病毒。表 9.1中列出了一些病毒的蛋白质种类数目和亚基的大致分子量以及所含核酸的种类。有的病毒还具有脂蛋白包膜(envelope)，它们的某些蛋白质常为糖蛋白，这些糖类通常是由岩藻糖、半乳糖、氨基葡萄糖以及甘露糖等组成的复杂多糖。此外，有些病毒还含有其他组分，例如在许多病毒中发现了多胺和金属离子如铁、钙、镁、铜、铝等。

9.1　病毒的形态学和整体结构

在电子显微镜下可观察到，有的病毒呈杆状或丝状颗粒；有的呈近球状颗粒（也称为球状或等轴颗粒）；还有些病毒是这两种形状的组合（例如一个噬菌体颗粒，具有球形的头部和杆状的尾巴）；有些病毒，特别是包膜病毒，大多是多形性的，即在某一群颗粒中既有球状的颗粒又有杆状的颗粒。图9.1示出了一些病毒的电子显微镜照片。

表9.1　一些病毒的蛋白质亚基和核酸种类

病　　毒	组成蛋白质的种数	亚基大致的分子量	核酸种类*
蚕豆斑驳病毒	1	20500	RNA
黄瓜病毒4号	1	15000	RNA
马铃薯 X 病毒	1	27000	RNA
芸豆南方花叶病毒	1	30000	RNA
烟草花叶病毒	1	17500	RNA
番茄丛矮病毒	1	41000	RNA
芜菁黄化花叶病毒	1	20000	RNA
腺病毒	9	7500—120000	DNA
大肠杆菌噬菌体 f2	2	14000和39000	RNA
大肠杆菌噬菌体 φX174	4	5000—48000	DNA-SS-c
大肠杆菌噬菌体 T4	28	11000—120000	DNA
脊髓灰质炎病毒	4	6000—35000	RNA
多型瘤病毒	6	13000—43000	DNA-c
猿猴病毒 40	3	16000—17000	DNA-c

* 除了另加注明外，所有 RNA 都是单链的(SS)，所有的 DNA 都是双链的(dS)，环状核酸用 c 表示。

为了讨论病毒的立体结构，首先要弄清一些有关的专用述语。病毒粒子(virion)指成熟的具有潜在侵染力的病毒颗粒；[病毒]壳体(capsid)指围绕着病毒核酸并与之紧密相连的蛋白质外壳；核壳(nucleocapsid)指壳体与核酸的结合体；结构单位(structure units)指壳体的蛋白

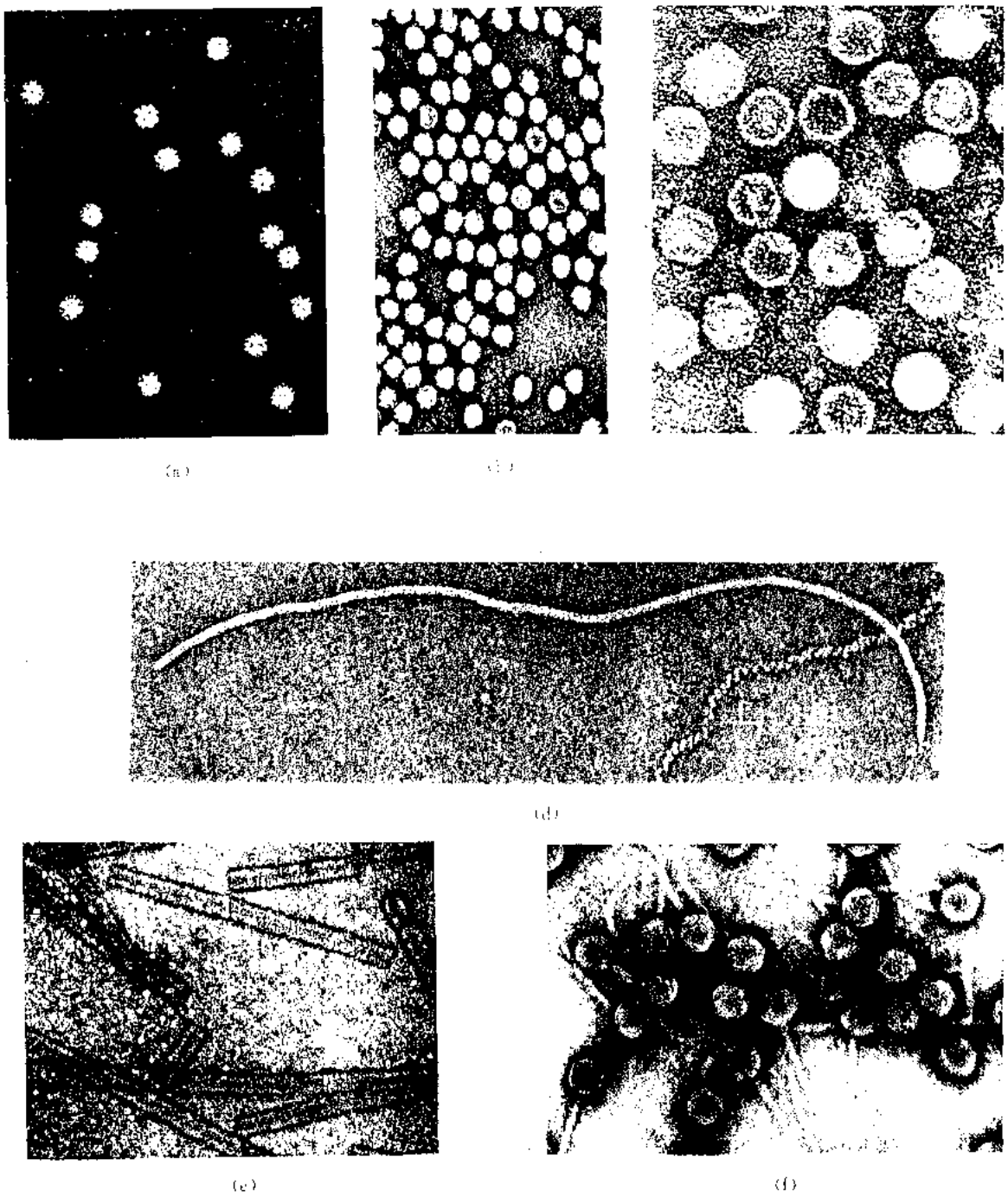


图9.1 病毒的电子显微镜照片(放大倍数不同)

(a),(b),(c)分别为球状病毒番茄丛矮病毒(TBSV)、芜菁黄花叶病毒(TYMV)以及脊髓灰质病毒(Polio);(d)为丝状的大肠杆菌噬菌体 fd;(e)为杆状的烟草花叶病毒(TMV);(f)为具有头部和尾部的噬菌体 λ

质亚基;[病毒]壳粒(capsomer 或 capsomere)或形态学单位(morphologic units)指在电子显

微镜下可以辨认的[病毒]壳体的亚结构，它可以是单个的结构单位，也可以是一簇结构单位；包膜(envolope)是由脂质、蛋白质、糖组成的结构，通常包被着在寄主细胞膜上的成熟的病毒粒子(virion)；刺突(spike)是指病毒表面的凸起的结构。

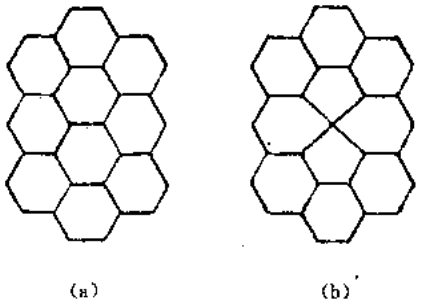


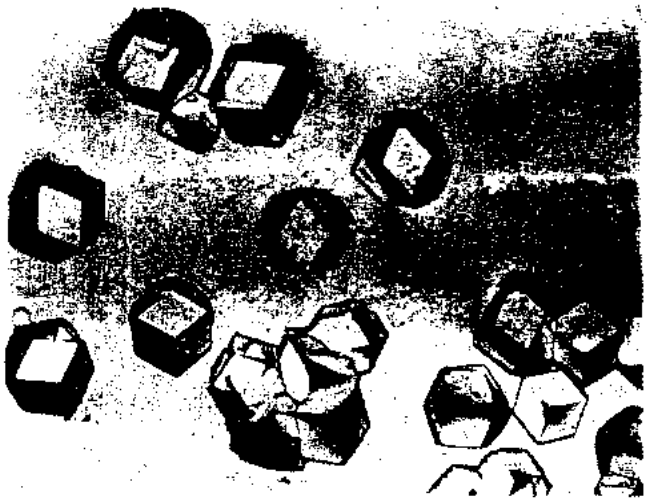
图9.2 两种成键排列方式
(a)等价键相连，其中所有内部键都是相等的；(b)准等价键，虽保持了整体结构，但内部键发生了变化

关于病毒颗粒整体结构的问题，F. H. C. Crick 和 J. D. Watson[88]根据各种物理的、化学的以及生物学的研究所提供的资料，于1957年提出了一个重要的解释和理论。他们认为，由于简单病毒的颗粒十分微小，它们仅仅含有一定限量的核酸，因而能从有限量的核酸密码所翻译成的病毒外壳蛋白质种类也是有限的。例如 TMV 的核酸 RNA 只有6400个碱基，按三联体密码原理计算，它至多也只能为10种蛋白质(按每种蛋白质由200个氨基酸残基组成计算)进行编码。后来对很多较简单的病毒外壳蛋白质进行的化学分析结果，都证实了这一推断，参考表9.1。正因为每个病毒颗粒的外壳蛋白质的种类有限，因而从理论上可以推断，它们在病毒核酸外面排列的方式也是很有限的。因为，一级结构相同的各个亚基一般具有相同的二级结构和三级结构，这些立体结构相同的亚基彼此会以相同的方式识别并以相同的各种性质的次级键结合在一起，而形成有规则的或对称的排列堆积。这是既经济又高效的病毒自我组装方式。也就是说病毒具有的规则性和对称性是生物学的基本要求。而要形成这样的规则而稳定的结构，至少要满足两个基本条件：第一，各亚基间形成最大数目的特异的化学键，将所有亚基有规律地等价地结合在一起；第二，这种结构必须处于自由能最低的状态，因而是最稳定的结构。不过，与一般晶体中相同的分子之间严格的等价结合有所不同，病毒外壳蛋白质的结构中，各亚基往往并不是严格等价地结合，而是所谓的准等价(quasi-equivalence)地结合，即在一个有规律的结构中，化学键等价相结合的方式有稍许而又不是紊乱的变化，从而使这种结构有可能比严格的等价结合更为稳定，见图9.2中的例子。根据现有的资料表明，绝大多数病毒分属于三种主要的对称组：螺旋对称性、二十面体对称性以及这两种对称性组合的复合型。

9.2 病毒的结晶和样品制备的一般原理

与脲酶被结晶揭示了酶的蛋白质化学本质一样，早在1935年 Wendell M. Stanley[89]首次报道了烟草花叶病毒(TMV)的类晶体，指出这类传染性的致病物是一种可以结晶的、分子量惊人地高达几百万的蛋白质(当时由于未检验制品中的磷，后来，人们证实了核酸也是病毒的成分)，揭示了病毒的化学本质，推翻了当时公认的所谓"病原说"理论。他用硫酸铵反复沉淀 TMV 提取液的方法，并用碱式醋酸铅脱色，获得了大量纯化的病毒。向这样的病毒悬液中加入足量的饱和硫酸铵产生混浊，然后逐渐加入50%饱和度硫酸铵的5%醋酸溶液，得到了类晶体。即使浓度稀释至10^{-9}g/ml 也仍有侵染性，而且即使经过多达10次的连续重结晶也不丧失其侵染性。制品的侵染性依赖于蛋白质的完整性。后来许多其他的植物病毒和动物病毒相继被结晶。例如1938年 J. D. Bernal 等人[90]报道了第一个植物球形病毒——番茄丛矮病毒(TBSV)

的晶体，1959年[91]报道了第一个动物球形病毒——脊髓灰质炎病毒(Poliovirus)的晶体。图9.3示出了这些病毒的晶体。



(a)

(b)

图9.3 球状病毒的晶体举例
(a)番茄丛矮病毒的晶体；(b)脊髓灰质炎病毒的晶体

完整的病毒颗粒和它的外壳蛋白质可以分别被纯化和结晶并分别进行X射线衍射单晶结构分析或纤维结构分析。与病毒整体分子相比较，外壳蛋白质分子量较小，较容易结晶和进行结构分析，因此某些病毒的外壳蛋白质结构的研究比病毒颗粒本身更为深入。病毒的外壳蛋白质各亚基之间以及它们与病毒内部的核酸之间是由各种次级键相结合的。因此，用一些能破坏氢键、盐键或疏水键等次级键，而又不至于破坏共价键的试剂进行处理，能将病毒颗粒中的

蛋白质与核酸等非蛋白质部分分开，或使蛋白质外壳解聚。常用的试剂为稀碱、稀酸、1—2mol/l 的各种盐类、高浓度蛋白质变性剂以及酚或十二烷基硫酸钠等。例如，在稀碱如 $NaHCO_3$、NaOH、硼酸盐或乙醇胺等在 pH10 -10.5下，即可将 TMV 颗粒的蛋白质部分释放出来；某些病毒如芜菁花叶病毒和 TMV 的一些株系，对碱性条件敏感，可用醋酸来分离和解聚；某些病毒如脊髓灰质炎病毒、马铃薯 X 病毒等，酸碱处理都不足以使蛋白质和核酸解离，需用高浓度的蛋白质变性剂如尿素、盐酸胍(2mol/l)等；又如苜蓿花叶病毒或雀麦花叶病毒等，则用高浓度的盐类如1—2mol/l 的 NaCl、$CaCO_3$进行处理等。上述各种方法，都有可能使病毒蛋白质不同程度地变性，有的是可逆变性，有的是不可逆变性，必须注意。

9.3 病毒的 X 射线衍射分析研究简述

关于病毒的 X 射线衍射研究始于19世纪30年代，特别是在 J. D. Bernal[92,93]的实验室中，早在1936年就用棒状的烟草花叶病毒(TMV)的定向排列的胶状样品拍摄了病毒的第一个纤维衍射花样(R. W. G. Wyckoff 等人[94]在此之前观察了 TMV 的 X 射线溶液散射)，并于1938年获得了球状的植物病毒 TBSV 的晶体[90]。但是，直到70年代末，G. Stubbs 等人[95]于1977年和 S. C. Harrison 等人[96]于1978年才分别由 TMV 的0.4nm 分辨率纤维衍射数据和 TBSV 的0.29nm 分辨率单晶衍射数据，获得了能被解释的电子密度图。这是由于病毒分子十分巨大而复杂，如 TBSV 病毒粒子分子量几乎为1000万道尔顿，晶体的晶胞边长达38.3nm，一个不对称单位中就含有60万道尔顿以上；而棒状病毒如 TMV 颗粒则未能生长出适合于结构分析用的任何晶体，它的结构仍不得不用纤维衍射的方法进行测定，因而丢失了许多信息。这两个病毒以及随后其他一些植物病毒和动物病毒结构的测定完全是在实验技术和理论研究迅速发展的基础上才实现的。首先，要测定病毒等这样一些大的组装体，尤其是对于病毒的单晶结构分析来说，必须有精确的收集衍射数据的技术，即用聚焦好的高强度的 X 射线才能分辨和记录下病毒所产生的衍射数据；第二，必须有能进行大量数据计算的计算机硬件和软件资源；第三，必须发展处理数据和分析结构的一些新理论和新方法。

用 X 射线衍射法对病毒进行研究所取得的成果，对于研究病毒的组装和拆卸、病毒的进化、病毒的免疫学和受体结合以及药物设计等都提供了有价值的资料。

9.4 具有螺旋对称性的病毒

这类病毒，由于组成它的蛋白质亚基是按照各种螺旋对称性堆积起来的，因而具有螺旋对称性，并形成棒状或丝状的病毒颗粒。螺旋对称性病毒包括棒状的植物病毒、丝状的噬菌体以及若干种动物和植物的有脂蛋白被膜包裹的病毒粒子。TMV 是这类病毒中最典型的、研究得最清楚的代表。见图9.1中 TMV 的电子显微镜照片。

9.4.1 螺旋对称性病毒完整颗粒的样品制备和 X 射线纤维衍射分析方法

为了得到十分好的纤维衍射照片以便对棒状或丝状病毒进行详细的研究，必须不断改进样品制备技术，使样品中病毒颗粒的排列精确地相互平行，标准偏差小于0.5°。为此，病毒颗粒的溶胶的浓度必须是足够高的(大约为15—30%)，以确保胶中颗粒定向好而且稳定，并可以减少水对射线的散射；但胶的粘度又必须低到足以使胶能自由流过直径为0.5—1.0mm 的毛细

管；pH 值必须远离等电点，以免颗粒间排斥力降低而使定向较困难，并防止颗粒间相互聚合；样品制成后必须放置足够的时间，以保证颗粒间的协同的相互作用，从而定向更好，有时甚至于原来定向不好的样品经放置一段时间后，定向得到明显的改进。另外，发现在制备纤维时，用强磁场也常能改进定向，例如丝状噬菌体的样品。现以 TMV 样品制备为例[97]：用磷酸缓冲液（或硝酸咪唑缓冲液），pH7.2（TMV 的等电点为 pH7.0），离子强度 $I=0.1$。将按照常规方法纯化了的 TMV 颗粒用缓冲液配制成浓度为25—30mg/ml 的病毒悬液，5900g 离心16小时，除去上层液体，将下层的溶胶吸入一个薄壁的石英（为减少对射线的吸收）毛细管中（直径约为 0.5mm），再吸入一些缓冲液，封管。用 CuK$_\alpha$ 细聚焦 X 射线管和铅玻璃毛细管准直器（直径 50μm）。在连续抽真空的照相机中，样品至底片的距离为12cm，用多层底片拍摄纤维衍射照片。图9.4为 TMV 的纤维衍射花样[98]，层线和斑点清晰，可以与单晶衍射媲美，而与前面两章

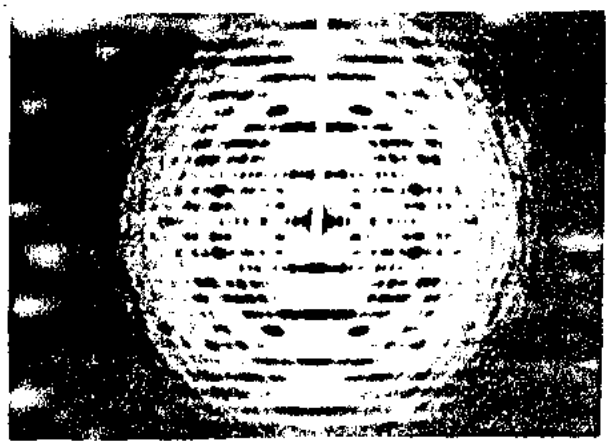


图9.4　TMV 定向溶胶的 X 射线纤维衍射照片

中讨论过的蛋白质或核酸的纤维衍射图中仅有数量很少的模糊的衍射斑点大不相同。因此，可以用类似于单晶衍射的方法计算出很好的电子密度图。不过，由于在病毒样品中，病毒颗粒的长轴虽然平行于毛细管的轴，但是它们是各自围绕着自己的长轴以无序的旋转角度而排列起来的，而且病毒颗粒不是严格地按点阵排列的。因此，病毒的纤维衍射图实际上也是记录了单个病毒颗粒的圆柱状平均了的结构因子的平方，因此丢失了一些信息，而且纤维图的柱状平均了的数据的相角问题也不同于单晶衍射的相角问题[99]。可将蛋白质单晶衍射分析中的重原子同晶置换法应用到病毒的纤维衍射分析中[98]，并用同晶置换法的多维类比的方法解决柱形平均的问题，计算出电子密度图。例如 TMV 颗粒[96]，制备了 OsO_4，Pb^{2+}，UO_2F_4，—Hg—CH_3 等多个重原子衍生物，计算出十分好的0.4nm 分辨率电子密度图，并用制约的最小二乘修正法扩充到0.29nm 分辨率[100]，能清楚地看到蛋白质的折叠以及 RNA 的构象及其与蛋白质的相互作用等，见图9.11。

9.4.2　外壳蛋白质的样品制备、晶体生长和结构分析

螺旋形病毒的外壳蛋白质可在一定的蛋白质浓度、pH、离子强度和温度等条件下，与核酸分离，并可解聚或重新聚合成不同大小的蛋白质亚基的聚合体，可用纤维衍射法或单晶衍射法研究它们的结构。例如 TMV 外壳蛋白质在低蛋白质浓度和 pH13时解聚成2.0S 的亚基单体；

在较高蛋白质浓度和 pH11时为4—4.6S 聚合体；在 pH7—9时为20—50S 圆盘状颗粒，其中20S 和30S 分别为2圈和3圈亚基的蛋白质盘(disk)；pH 低于6.3时，重新聚合成长棒形完整的蛋白质外壳螺旋体。图9.5为 TMV 外壳蛋白质的各种聚合体的示意图。图9.6为 TMV 外壳蛋白质20S 圆盘的电子显微镜照片。

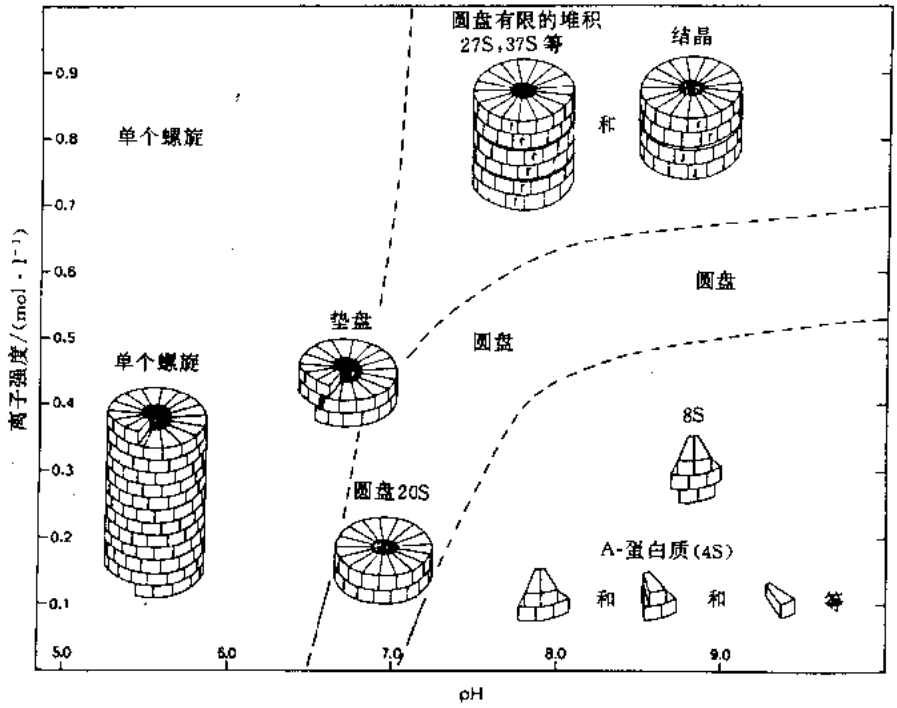


图9.5　TMV 外壳蛋白质的各种聚合体示意图

其中 A-蛋白质是沉降系数小于20S 的低分子量的寡聚体的混合物，其主要成分为4S 寡聚体；垫盘是 TMV 组装过程中的一个暂时的中间物，20S 圆盘通过垫盘形成螺旋

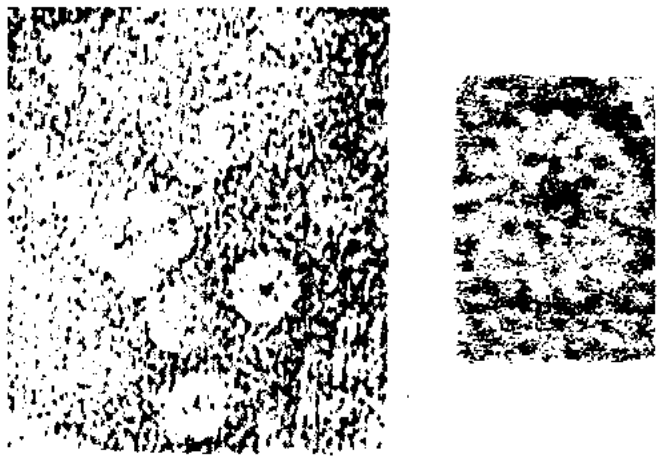


图9.6　TMV 外壳蛋白质的20S 盘的电镜照片

TMV 外壳蛋白质的单晶 X 射线衍射研究也是十分成功的例子。1974年，A. Klug 的实验室[101]对 TMV 亚基的小聚合体进行了系统的结晶实验，得到了20S 圆盘的晶体。20S 圆盘的制

备和结晶方法为：1%TMV 病毒悬液在4℃下对0.1mol/l 乙醇胺-HCl，pH10.5的缓冲液彻底透析2—3天，再对水透析1天。然后在室温下对 Tris-HCl，pH8，$I=0.1$的缓冲液透析1天。将此降解了的病毒悬液在140000g 下离心75分钟，除去 RNA 沉淀。上清液与等体积的3mol/l 硫酸铵混合，静置10分钟。在12000g 下离心20分钟，回收蛋白质沉淀。按初始病毒悬液中病毒的含量每30mg 加入1ml Tris-HCl，pH8，$I=0.1$的缓冲液到沉淀中，使之溶解。在16000g 下离心30分钟，取上清液供结晶用。最佳晶体生长条件为：将含有8mg/ml 蛋白质、0.3mol/l 硫酸铵的 Tris-HCl，pH8，$I=0.1$缓冲溶液，置于室温下2—3星期，或者预先在15—16℃下放置过夜，即能得到较好的晶体。晶体为正交晶系，空间群为 $P2\ 2_12$，$a=22.8$nm，$b=22.4$nm，$c=17.4$nm。A. Klug 的实验室[102]在1978年报道了利用同晶置换法和非晶体学对称性解出的0.28nm 分辨率 TMV 的圆盘的结构分析结果。

9.4.3 棒状植物病毒 TMV 的立体结构

J. D. Bernal 等人于[93]1941年对 TMV 的 X 射线纤维衍射进行的开拓性工作，为后来详细地研究 TMV 的结构铺平了道路。R. Franklin 等[98]于1958年将分别由病毒的完整颗粒与仅由亚基重新聚合成的完整蛋白质外壳的 X 射线纤维衍射数据所计算出来的径向电子密度分布曲线进行比较，发现前者在离颗粒的长轴约为4.0nm 的距离处有一个电子密度最大值，而在后者，此值消失了，见图9.7。因而很明显，这个最大值处即为 RNA 的位置。从图中还可以看到，TMV 颗粒的半径约为8.5nm，在内部是距长轴半径约为2.0nm 的空心通道。因而远在用电子显微镜直接观察以前，TMV 病毒颗粒的一个结构模型就已经从 X 射线衍射分析和生物化学分析的结果而建立起来了，见图9.8[103]。

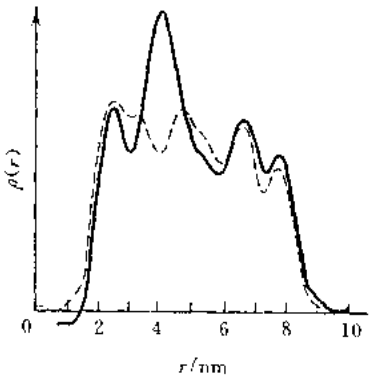


图9.7 TMV 完整颗粒(实线)和完整蛋白质外壳(虚线)的径向电子密度分布曲线

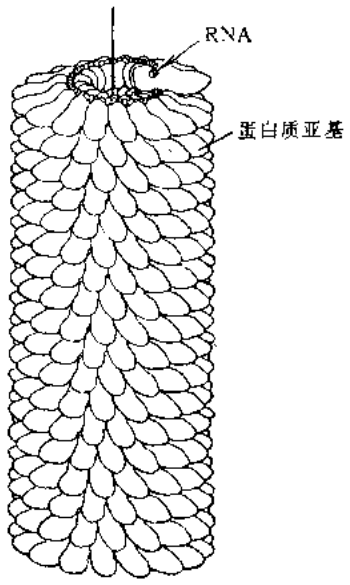


图9.8 TMV 的结构模型

G. Stubbs 等人[100]对 TMV 进行的 X 射线纤维衍射精确分析以及 A. Klug[101]实验室对 TMV 蛋白质盘进行的单晶 X 射线衍射结构测定，已在分子和原子水平上阐明了 TMV 蛋白

质亚基的肽链折叠、RNA 的构象以及两者的相互作用。

综合 X 射线衍射和生物化学分析以及电子显微镜观察结果表明，TMV 为棒状，长约300nm，直径约17nm，中心为一个直径约4nm 的空心通道。TMV 共有约2140个蛋白质亚基，它们围绕着病毒颗粒的轴成右手螺旋排列。所有亚基都是相同的，每个亚基的分子量为17420道尔顿。每3圈螺旋有49个亚基，螺距为2.3nm，轴向重复周期(3圈)为6.9nm。单链的 RNA 在离病毒颗粒的轴约4nm 处，与蛋白质结合并跟随着蛋白质亚基绕轴成右手螺旋盘绕，核苷酸之间的轴向距离约为0.51nm。

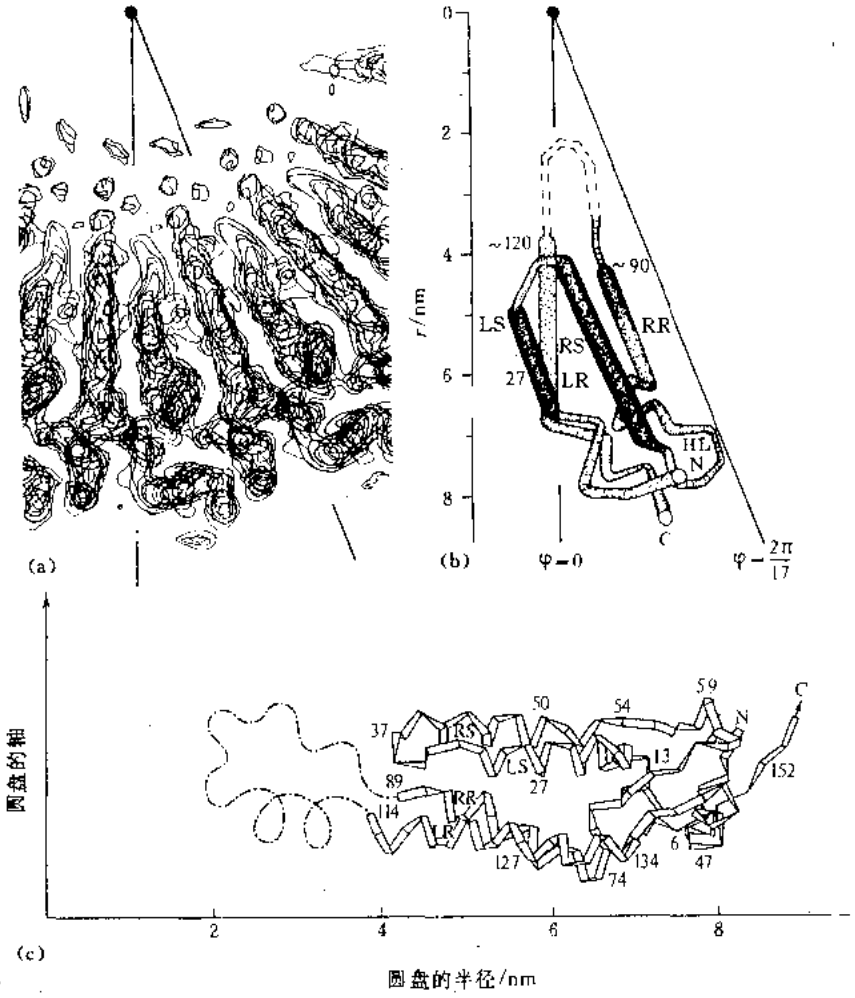


图9.9 TMV 中蛋白质亚基的结构和排列

(a)完整蛋白质外壳螺旋体的横截面电子密度图；(b)每个亚基径向排列的示意图；(c)每个亚基的结构

高分辨率的病毒颗粒的纤维衍射和蛋白质盘的单晶衍射数据的精确分析表明，在 TMV 蛋白质外壳的组装中的中间物是 TMV 的蛋白质盘。每个盘是由两层环组成的，每层有17个蛋白质亚基，见图9.9和图9.10。每个亚基的多肽链主要由折叠成近乎相互平行的紧密堆积的4个 α 螺旋组成核心，内部的氨基酸侧链形成疏水层，使蛋白质折叠稳定；外部为亲水性与疏水性

混合的特征外层，与其他亚基相互作用，使亚基堆积十分稳定；此外，还有两个另外的短螺旋片段以及连接4个核心螺旋的β折叠的一个小区域。由图中可以看到，TMV的亚基堆积成螺旋

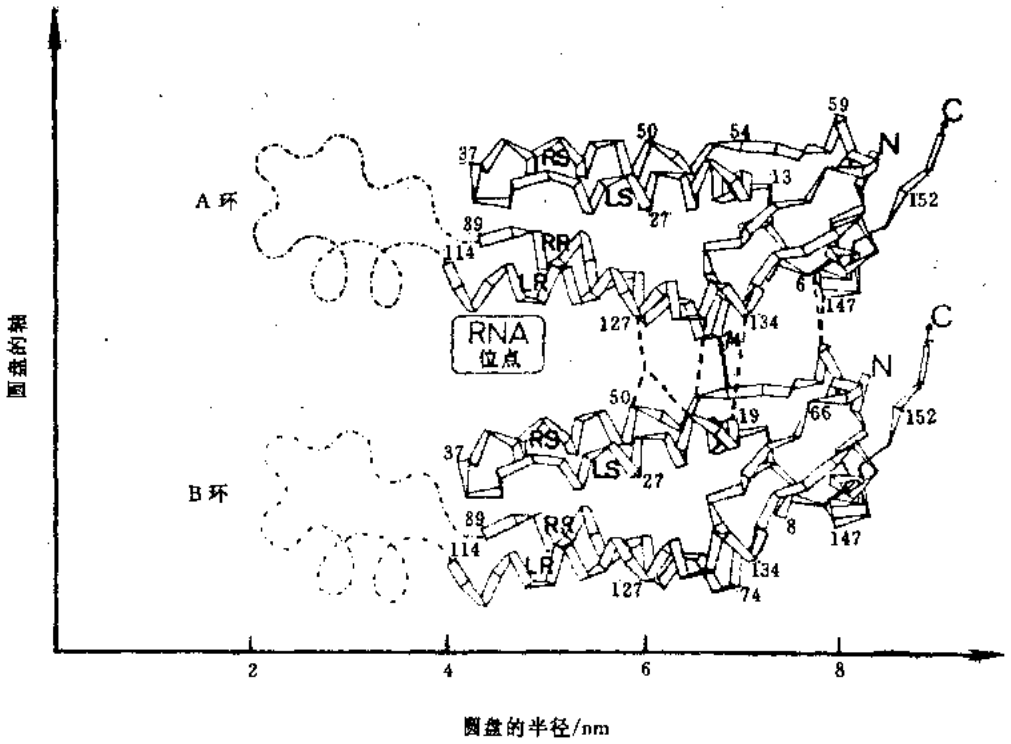


图9.10 TMV蛋白质盘的侧面观

表明了盘中亚基之间轴向的接触。虚线表示氢键结合，实线表示疏水接触

形是由亚基本身的形状(即朝着螺旋轴的方向逐渐变窄)以及亚基表面上的活性基团的配置而预先决定了的，因而保证了病毒的自我组装。首先是组装成蛋白质盘，然后组装成一个稳定的螺旋结构。而且，由图中可以看到，蛋白质盘的两层中，上下亚基之间仅在高半径处相互接触；而在低半径处即朝向盘的中央孔的一端，两层是分开的，正好容纳RNA并与RNA相互作用。

TMV的亚基界面处有2个电荷簇，它们可能是控制病毒的组装和拆卸的。一个簇是在低半径处的侧-侧界面上，另一个是在半径约为6nm处的上-下界面中。这两个电荷簇是由部分地被邻近Arg所中和了的羧基形成的，是结合Ca^{2+}和H^+的位置。Asp 116与一个磷酸根十分接近，所产生的静电排斥可能在病毒的拆卸中是重要的。

TMV中的RNA是十分有序的结构，见图9.11。每3个核苷酸残基与1个蛋白质亚基结合，TMV是至今已看到了蛋白质-核酸相互作用的唯一的病毒。RNA中的磷酸根被Arg^{41}，$Arg^{[illegible]}$以及Arg^{92}所中和，但是它们并不都形成简单的离子对；Asp^{115}，Asp^{116}以及一个磷酸根的负电荷被Arg^{112}部分地中和了；所有3个碱基都对着LR螺旋平置着，形成了疏水相互作用，这种作用是RNA序列非特异性的，其中2个碱基(1个为顺式构象，1个为反式构象)堆积在一起并竖直朝上指向亚基间盐桥Arg^{113}-Asp^{115}与Arg^{122}-Asp^{38}之间的一个空穴中，第3个碱基(反式构象)则在LR螺旋与下面一个亚基的连接LS与RS的环之间，沿着LR螺旋平置着；在RNA

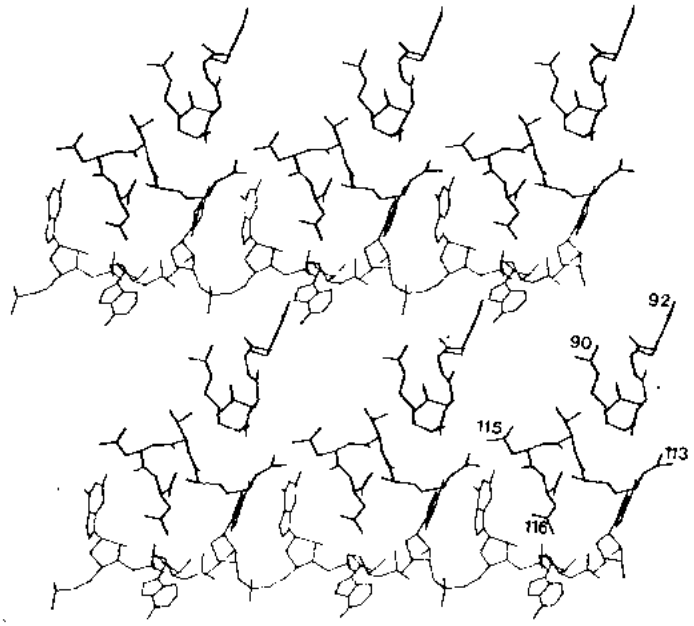


图9.11 TMV 中蛋白质亚基与 RNA 的相互作用

粗线为6个亚基的残基片段90－92和113－116；细线为 RNA 链。为沿半径朝轴方向的内部纵面观。RNA 的两个碱基竖直地伸展到两个相等的 α 螺旋之间的亚基间的空间中，另一个碱基沿着 α 螺旋伸向读者

上方和下方的亚基的侧链为这些碱基的极性部分提供了一个亲水环境。虽然大多数蛋白质-核酸相互作用不具有碱基特异性，但是可能有几个氢键是特异地有利于 AAG 序列的。已经知道这个序列对于病毒组装的早期阶段在识别作用中是重要的，而且 AAG 作为一个三核苷酸，牢固地结合到 TMV 蛋白质上。

TMV 许多年来已经成为 X 射线纤维衍射和病毒结构研究的一个模型系统。到现在为止，虽然对 TMV 以外的螺旋对称的棒状病毒尚无详细的结构测定资料，但是已报道了若干种这类病毒的初步 X 射线衍射结果，其中包括 tobamoviruses 族（TMV 也属于这个家族）、potexviruses 族、tobaccorattle virus 族以及 barley stripe mosaic virus 族中的若干种病毒。

9.4.4 丝状的噬菌体 Pf1的立体结构[104]

噬菌体 Pf1是除了 TMV 以外的至今已在分子细节上测定了空间结构的唯一的螺旋对称形病毒。它是长约1 940nm，直径约6nm 的柔韧的丝状噬菌体。它含有一个单股 DNA 分子和几种不同的外壳蛋白质亚基。

L. Makowski[105]和 D. A. Marvin 等人[106]分别发表了高温型式的0.7nm 分辨率和低温型式的0.3nm 分辨率的 Pf1纤维衍射研究结果。其 X 射线纤维衍射花样见图9.12。结构分析表明，其中数千个一种主要的亚基沿着噬菌体的长轴方向螺旋对称地排列着，见图9.13。螺旋的精确对称性依赖于温度、重原子的结合以及其他因素。螺旋的每一圈有5.40至5.46个亚基不等，若干种次要的外壳蛋白质亚基则组装在噬菌体丝状颗粒的末端处。DNA 包装在蛋白质外壳内。固态 NMR 实验表明 Pf1中的 DNA 是有序的结构，但现在尚无 X 射线衍射分析的数据。

图9.12　Pf1的X射线纤维衍射照片

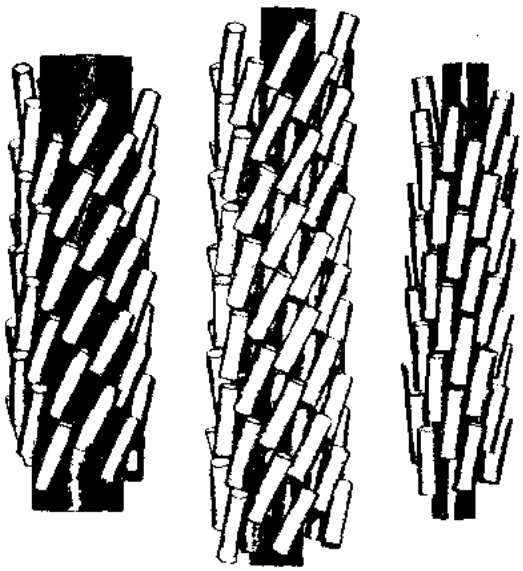


图9.13　在Pf1中两层α螺旋排列的示意图(中图)
中央的深色大柱体代表被DNA占据的直径约为2nm的中央核心；
小柱体代表长约3nm，直径约0.9nm的α螺旋。为了更清楚可见，
将两层α螺旋分别表示在右图和左图中

在图9.13的模型中，每个主要的蛋白质亚基分子含有两段直的α螺旋。一段靠近轴，与病毒长轴成6°夹角；另一段在病毒表面，与轴成26°，因而形成双层的α螺旋排列。亚基中链的走向是从外部α螺旋段的N端到内部α螺旋段的C端。两个α螺旋段之间的连接尚未确定，可能是被破坏了的某种二级结构，也可能是一个β转角，也可能不是两段α螺旋而是一个连续的

弯曲的螺旋。

Pf1是在分子水平上已测定了空间结构的第一个丝状病毒。此外，还有几个丝状病毒也在不同水平上进行了X射线衍射研究，例如噬菌体Xf，fd，MS2以及M13等。

9.5 球形病毒

球形病毒包括许多动物病毒和植物病毒。这些病毒的外壳蛋白质亚基排列成近似于球形，核酸位于紧密包裹的蛋白质的球形外壳中。更精确地说，亚基或衣粒以二十面体的对称性排列成球形的外壳。由于它们的结构规则并近似于球形，所以许多这样的病毒能形成很好的晶体，并已用X射线单晶衍射方法进行了详细的研究。

9.5.1 二十面体的几何学

二十面体是体积-表面积比最有利的一种多面体。一个规则的二十面体有20个等边三角形

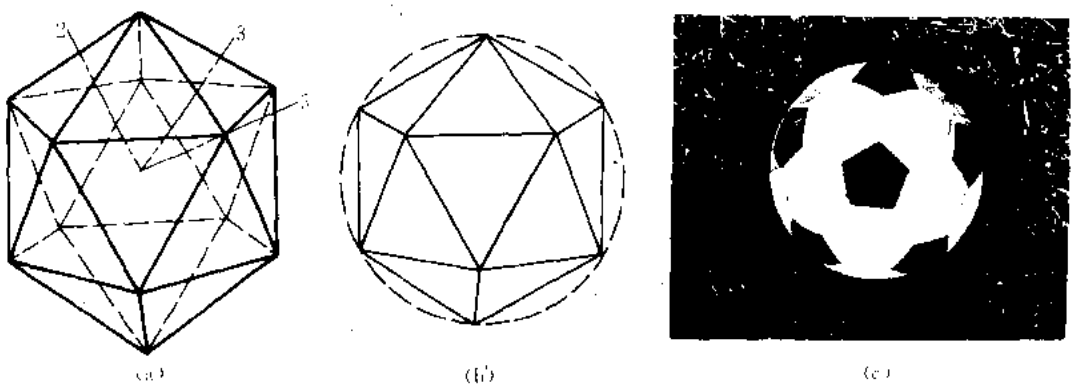


图9.14 二十面体

(a) 一个规则的二十面体及其二重、三重和五重旋转轴对称性；(b) 规则二十面体近似的球形表面；(c) 足球的二十面体对称性

的面，12个顶点和30条边，见图9.14。当从每个顶点观看时，有五重旋转轴对称性；当从每条边的中点垂直观看时，有二重旋转轴对称性；当从每个正三角形面的中心垂直观看时，有三重旋转轴对称性。也就是说，二十面体具有5·3·2对称性。二十面体内接于一个球中，并接近于球形，其顶点与球面接触。

不仅规则的二十面体具有5·3·2对称性，而且由粗略的三角形和六角形所构成的二十面体也具有这种对称性，一个球形物体例如足球也可具有同样的对称轴排列。

9.5.2 球形病毒的二十面体对称性

D. L. D. Caspar 和 K. Klug[107]早在1962年就对球形病毒的结构规律进行了理论推导，指出可将大多数小的球形病毒看成是一系列理想的二十面体结构。为了表明各种不同的球形病毒的形态单位(即壳粒)的总数与结构单位(即亚基)的总数之间的关系，并借以描述各种球形病毒的外壳结构，他们提出了一个“三角形剖分数”(triangulation numbers)的理论。将二十面体的每个等边三角形面划分成 T 个单位等边三角形，由于每个二十面体有20个面，所以单位三角形的总数是$20T$，因此，可以用“三角形剖分数”T 值表示球形病毒的结构类型(图9.15)。

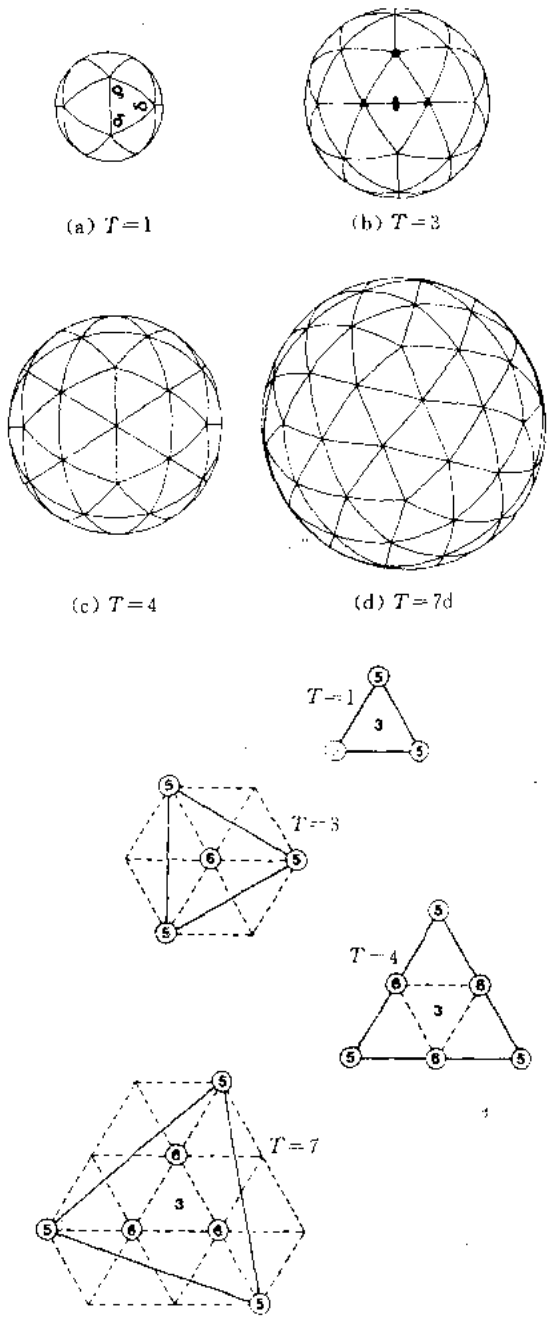


图9.15　二十面体被划分成三角形的表面排列的常见类型

8符号代表蛋白质亚基，3表示三重轴，5表示五重轴，6表示局部的六重轴

最有可能的即最小的4个 T 值为1，3，4和7。由于一个蛋白质分子或亚基本身的结构是不对称的，所以每个三角形中至少有三个亚基。也就是说，球形病毒的蛋白质外壳中能容纳60T 个亚基。由图中还可看到，正是由于二十面体的三角形表面能再分成单位三角形，因此，除了二十面

体本身所固有的五重旋转轴、三重旋转轴和二重旋转轴对称性外，还产生了局部的六重旋转轴对称性，因而具有围绕着五重旋转轴的五位体(penton)和围绕着六重旋转轴的六位体(hexon)两种形态学单位即壳粒。T 值与壳粒数目的关系见表9.2。这两种壳粒具有几乎相同的亚基间相互作用，可能只要求亚基间的次级键稍许变化即可。虽然这些亚基的结构不是严格地相等的，但可以说是准等的。

表9.2 T 值与壳粒数的关系

T 值	单位三角形总数	五位体壳粒数	六位体壳粒数	壳粒总数	实　例
1	20	12	0	12	φX174噬菌体
3	60	12	20	32	芜菁黄花叶病毒
7	140	12	60	72	多形瘤/乳头状瘤病毒
16	320	12	150	162	疱疹病毒
25	500	12	240	252	腺病毒

T 值的计算公式为 $T=H^2+K^2+HK$，H 和 K 为任意正整数，包括零在内。在此，H 和 K 是以从一个五重轴开始经过局部的六重轴到相邻的另一个五重轴所走过的步数来计算的，转折前和转折后的步数分别定义为 H 和 K 的值，并规定 $H>K$。例如图9.15中的几个常见的对称性类型：若从一个五重轴直接可到相邻的另一个五重轴，无六重轴，则 $H=1,K=0,T=1$；若从一个五重轴，经过一个六重轴后转折，再到另一个相邻的五重轴，则 $H=1,K=1,T=3$；若从一个五重轴，经过一个六重轴，不经转折再到另一个相邻的五重轴，则 $H=2,K=0,T=4$；若从一个五重轴，经过两个六重轴后转折，再到相邻的另一个五重轴，则 $H=2,K=1,T=7$。

当然，用这种理论的人工构造来作为球形病毒外壳的结构模型是太有限了。首先，由于自然界具有极大的复杂性和多样性，某些球形病毒可能具有另外的排列类型；而且，必须指出，"三角形剖分数"并不能代表球形病毒自我组装的实际性质。一些小病毒是按"全或无"原则形成的；较大的二十面体病毒是逐步组装成的，例如腺病毒是先形成外壳的各个面，而且可以独立存在，然后这些面再结合形成类似于二十面体的外壳；还有一些病毒是先形成一个小的内部的蛋白质囊，核酸结合于其上，然后再形成外部的蛋白质壳。但是，"三角形剖分数"理论提出后30多年来，指导人们了解了病毒中亚基的堆积原理，并且许多概念已经被结构测定所证实和应用。

例如，$T=1$的植物病毒中已解出晶体结构的如卫星烟草枯斑病毒(STNV，即 satellite tobacco necrosis virus)[108]，直径约17nm，含有一个单股的 RNA 分子和60个化学上相同的蛋白质亚基，即二十面体的每个三角形面上有3个亚基。

又例如，现已测定了晶体结构的球形病毒多为 $T=3$，直径约30nm 以上，内含 RNA 的病毒，如植物病毒中的番茄丛矮病毒(TBSV，即 tomato bushy stunt virus)[95]、芸豆花叶病毒(SBMV，即 southern bean mosaic virus)[109]等。它们的蛋白质外壳均各由180个化学上相同的亚基(或称为原体 protomer)所组成，二十面体的每个三角形面上再划分为3个单位三角形，每个单位三角形中有3个亚基 VP1，VP2和 VP3。并以12个五位体和20个六位体的准对称性形成球形的外壳，即共有32个形态学单位——壳粒；又例如动物病毒中的人感冒病毒14(HRV14，即 human rhinovirus 14)[110]、脊髓灰质炎病毒(Polio，poliovirus type 1)[111]和 Mengo 病毒[112]

等，各含有4种蛋白质亚基（原体）VP1，VP2，VP3和 VP4，每种亚基有60个拷贝，其中 VP1，VP2和 VP3的结构是相似的，而且与上述 $T=3$的植物病毒的蛋白质亚基也是相似的。PV4位于病毒的内表面，因而这4种亚基仍以12个五位体和20个六位体的准对称性组装成蛋白质外壳，壳粒数目也是32。

又例如 $T=25$的腺病毒（adenovirus）[113]，直径约90nm，比前述所有病毒都大得多。它是一种含有 DNA 的动物病毒。它具有十多种不同的蛋白质亚基，组成12个五位体复合物和240个六位体，即在外壳表面上共有252个壳粒。

9.5.3 球形病毒的晶体生长和结构分析

由于球状病毒的结构规则并近似于球形，所以许多这样的病毒能形成很好的晶体，尤其是小的球形病毒。因此，可以用单晶衍射方法直接研究球形病毒颗粒。我们知道，病毒颗粒的外部为蛋白质外壳，因此，球形病毒颗粒的结晶方法与蛋白质相似。例如，第一个被解出晶体结构的球形病毒 TBSV[96,114]，是在10mg/ml 病毒悬液中，加入饱和硫酸铵至0.8mol/l 于4℃下放置数月后，生长出了适合于 X 射线衍射研究用的晶体。为了增加晶体的稳定性，可小心地将晶体转移到4mol/l 的硫酸铵溶液中，或1mol/l 的硫酸钠溶液中或含0.8mol/l 硫酸铵以及0.5g/ml蔗糖的溶液中。TBSV 的晶体见图9.3，它的旋进法衍射照片见图9.16。

用球形病毒颗粒的晶体进行 X 射线单晶衍射结构分析，已得到了很好的结果。可以用常规的重原子同晶置换法，并利用二十面体对称性原理等计算相角和电子密度图。但是，病毒颗粒比蛋白质分子大得多，晶胞相当大，因而衍射数据也相当多，所以病毒的晶体结构分析工作比蛋白质更为复杂和困难。例如 TBSV 为立方晶系，空间群为 I23，$a=38.32$nm。一个晶胞中含有2个直径为33nm 的病毒颗粒。收集了母体和2个重原子衍生物的0.29nm 分辨率反射各约20万条，用非晶体学的二十面体对称性进行几轮对称性平均，然后用多对同晶置换法计算相角，最后计算出电子密度图。

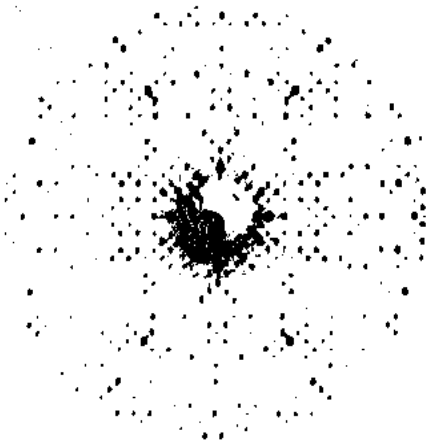


图9.16 TBSV 晶体的旋进法衍射照片

9.5.4 球形病毒 TBSV 的立体结构

如前所述，番茄丛矮病毒（TBSV）是第一个被结晶的并且第一个被分析晶体结构的球状病毒。我们以它的0.29nm 分辨率结构分析结果为例，了解球状病毒的立体结构。

TBSV 病毒外壳是由180个化学上相同的亚基组成的。这些亚基以 $T=3$的二十面体对称性在离颗粒中心半径为10.5nm至14.5nm 之间紧密地堆积成病毒的表面壳层，并延伸出90个突起至半径为17nm 处，见图9.17。

每个单位三角形中有3个亚基，它们以一个准三重轴（3^q）联系起来。按照它们与准五重轴（5^q）的距离从近到远，分别叫做 A、B 和 C。每个亚基由3个结构域组成：R 结构域（random domain）在氨基末端；其后是 S 结构域（shell domain）；P 结构域（protruding domain）在羧基末端。其中 S 结构域为楔形，因而窄的一端可以十分好地朝着准五重轴和准六重轴而紧密地堆积在一起，形成病毒的表面。P 结构域伸出表面并成对地组合在一起，S 与 P 之间由铰链（hinge）

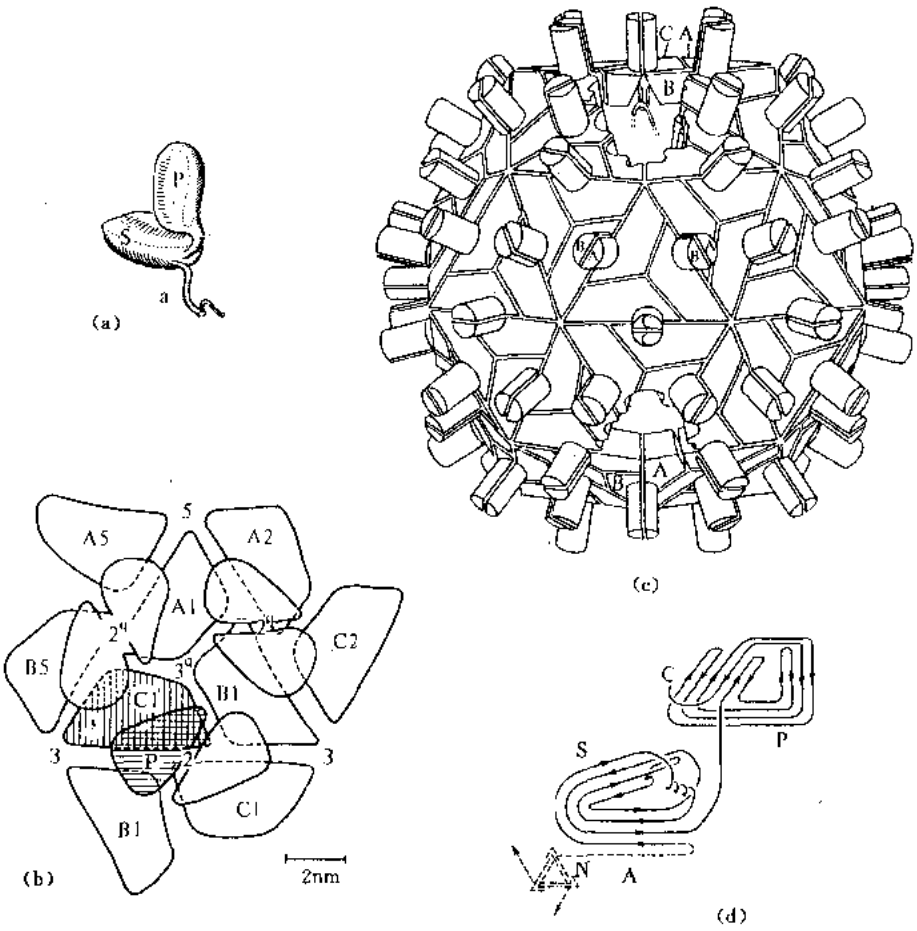


图9.17　TBSV 的结构示意图

连接。这两个结构域分别为8股的和10股的一个反平行 β 桶结构，见图9.18；它们的拓扑图称为“果酱卷”(jellyroll)，见图9.19，它是由两条连续的反平行的 β 链绕成一个螺圈；因此这种 β 桶又称为果酱卷 β 桶(jellyroll β barrel)。在 S 结构域中，C 链与 D 链之间以及 E 链与 F 链之间各有一个短的 α 螺旋。关于 R 结构域的结构，在准等的三个亚基中有所不同。在 C 亚基中，R 结构域为一条明显的 β 链，它与另外两个 C 亚基中的同样的 β 链形成一个围绕着准六重轴的由三条短链(3个残基)组成的 β 片层，又称为 β 环(β annulus)，见图9.18(c)；但在 A 和 B 亚基中，R 结构域不能清楚地被鉴别，而且深深地穿入到病毒颗粒内，它们的功能很可能是将蛋白质外壳与位于病毒内部的 RNA 联系起来。另外，RNA 在电子密度图中也未能看到。

关于亚基间的相互联系，在5^q，6^q 和3^q 周围亚基的接触是广泛的；但除了 P 结构域以外，在2^q 周围则是十分稀疏的。

此外，在 S 亚基表面上有一些电荷丛，其中3个在3^q 界面处，2个在5^q 界面处。钙离子结合在这些位点上。这些电荷丛对于病毒的组装和拆卸是十分重要的，很可能由于羧酸的电荷丛或羧酸和组氨酸的电荷丛的 pK 值接近中性，它们对环境因子中的电荷如 pH 或钙离子浓度等十分敏感，因而特别适合于在生理条件下起着静电开关的作用，以控制病毒的组装和拆卸。

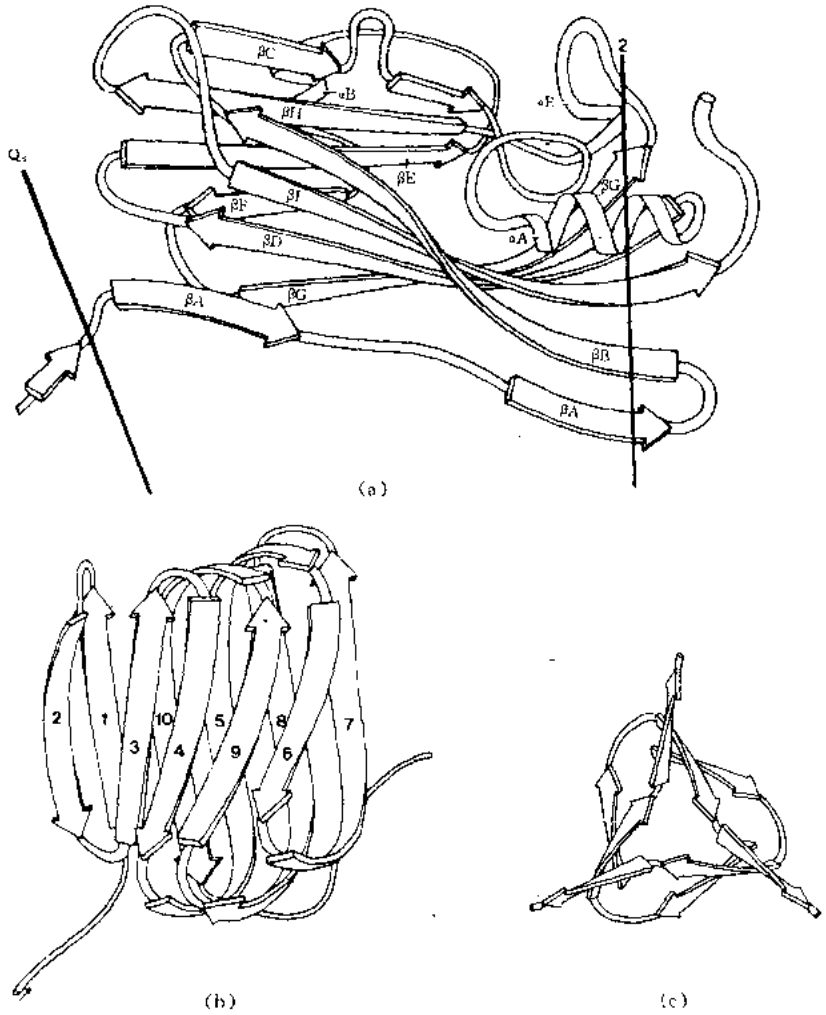


图9.18 TBSV 的 S 结构域(a)、P 结构域
(b)和 β 环(c)的结构(比例不等)

9.5.5 球形病毒外壳蛋白质肽链的折叠特点[115,116]

当十多年前首先测定出0.4nm 分辨率的 TMV 和0.29nm 分辨率 TBSV 的结构时，发现它们的外壳蛋白质的二级结构和三级结构是十分不同的。如前所述，前者主要是 α 螺旋，后者是 β 折叠。因此当时人们以为，对更多的病毒进行结构分析后一定会得到各种不同特点的相互无关的结构类型。但是，事实上自那以后所测定的病毒结构虽然有相当大的变化，但是令人惊奇的是，它们的结构类型的数目是极其有限的。直到现在为止，已用 X 射线晶体学方法测定了结构的所有球状病毒，甚至于在它们的氨基酸序列、血清学以及寄主范围等方面并不具有相似性的情况下，都已表明它们的结构是与 TBSV 这个最早的例子有关的。现在已经知道的病毒外壳蛋白质的结构，共有三种基本的折叠方式。除了前面已经讨论过的在棒状病毒如 TMV 中的 α 螺旋捆和丝状噬菌体如 Pf1中的相互分离的 α 螺旋两种方式以外，第三种折叠方式即为球形病毒的 β 桶结构。当然这并不是说不可能再发现其他类型，但是这确实表明了这些病毒的进

化关系和适应性。

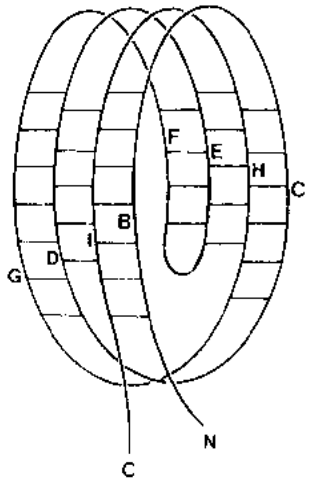
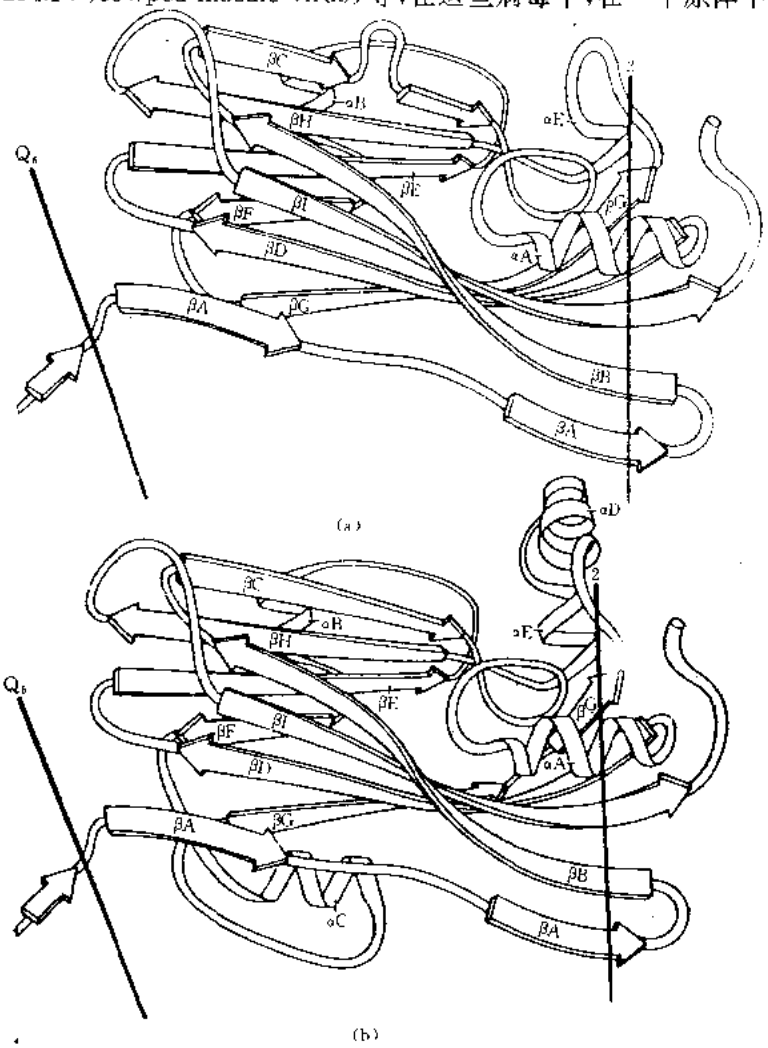


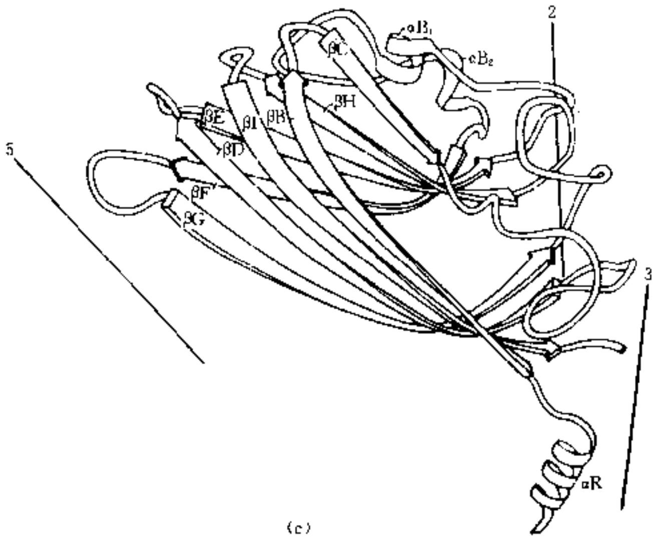
图9.19 "果酱卷"拓扑图

图9.20比较了球形植物病毒 TBSV，SBMV 的 S 结构域、STNV 的亚基以及球形动物病毒 HRV14，Mengo virus 以及 poliovirus 的 VP1，VP2和 VP3的结构。虽然来源很不相同，而且"三角形数目"的 T 值相同或不同，但是它们的外壳蛋白质亚基或亚基中的某些结构域却具有极其相似的结构。它们最共同的折叠特点是具有"果酱卷"β 桶、β 环以及两个短的 α 螺旋，正如图9.18中所看到的 TBSV 各个结构域那样，都极为相似。

图9.21为腺病毒的六位体亚基的结构。在这类病毒中，可以看到球形病毒外壳蛋白质结构的复杂性和多样性，但是它们同样也具有上述的折叠特点。图中可以看到，在底部的结构域 P1和 P2为两个"果酱卷"β 桶；上部的"塔"称为 T 结构域，它与另外两个亚基的相应的各自的2条 β 链形成"β 压缩"("β constriction")，是一个6股的管子，与在许多小的二十面体病毒中所发现的3股或5股的 β 环相似。属于这种折叠类型的已经研究过的病毒还有豇豆花叶病毒(CPMV，cowpea mosaic virus)等，在这些病毒中，在一个原体中有多个 β 桶。

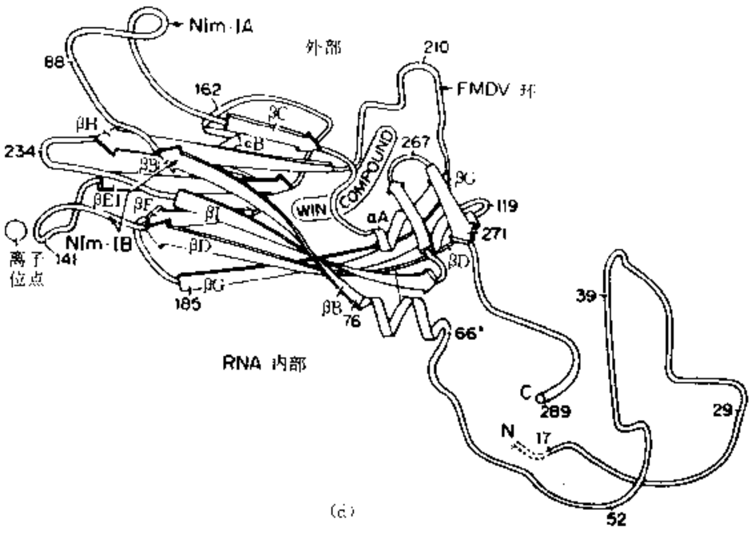


❸



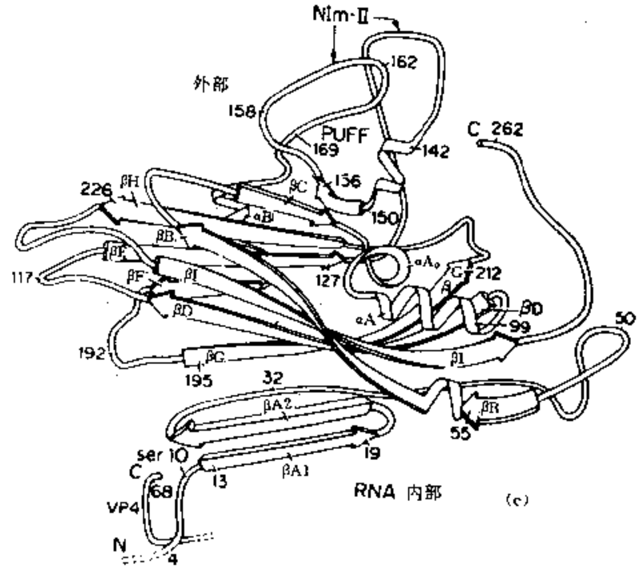
2
5
3
βC
βH
βE
βI
βB
βD
βF
βG
αR

(c)

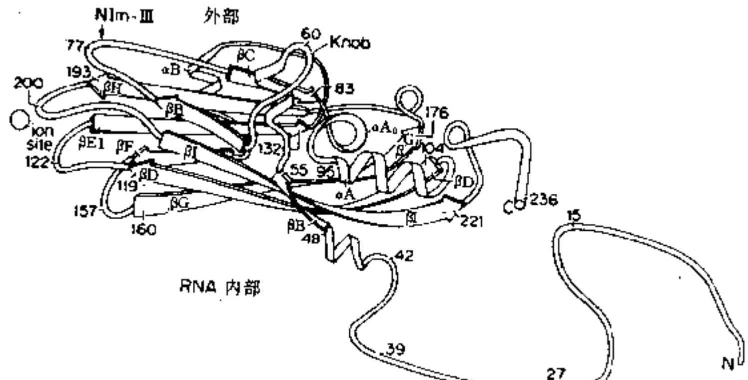


Nim-IA
外部
88
210
FMDV 环
162
βC
βH
αB
234
267
COMPOUND
WIN
βG
119
αA
271
Nim-IB
离子
位点
141
βD
βG
185
βB
76
66'
39
RNA 内部
C
289
29
N
17
52

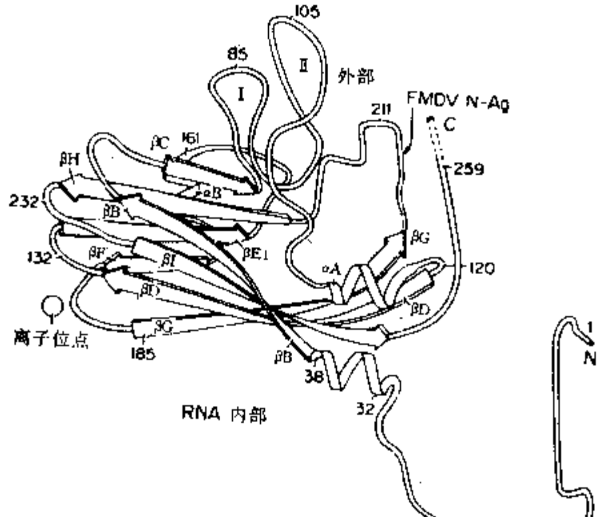
(d)


NIm-II
外部
162
158
PUFF
169
142
C 262
136
226
βH
βC
150
αB
βB
βE
αA₀
212
117
βF
βI
127
βD
αA
βD
99
50
192
βG
βI
195
32
βA2
βB
55
ser 10
19
C
13
βA1
68
RNA 内部
VP4
N
4

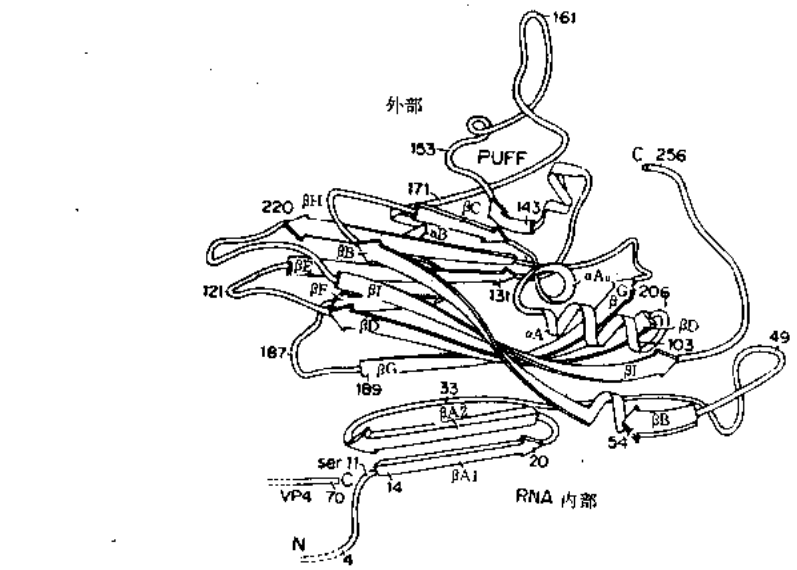
(c)


NIm-III
外部
60
Knob
77
βC
193
αB
83
200
βH
βB
176
ion site
αA₀
βE1
βF
βI
132
104
122
55
95
βD
119
βD
αA
157
βG
236
160
βI
221
15
βB
48
42
RNA 内部
39
27
N

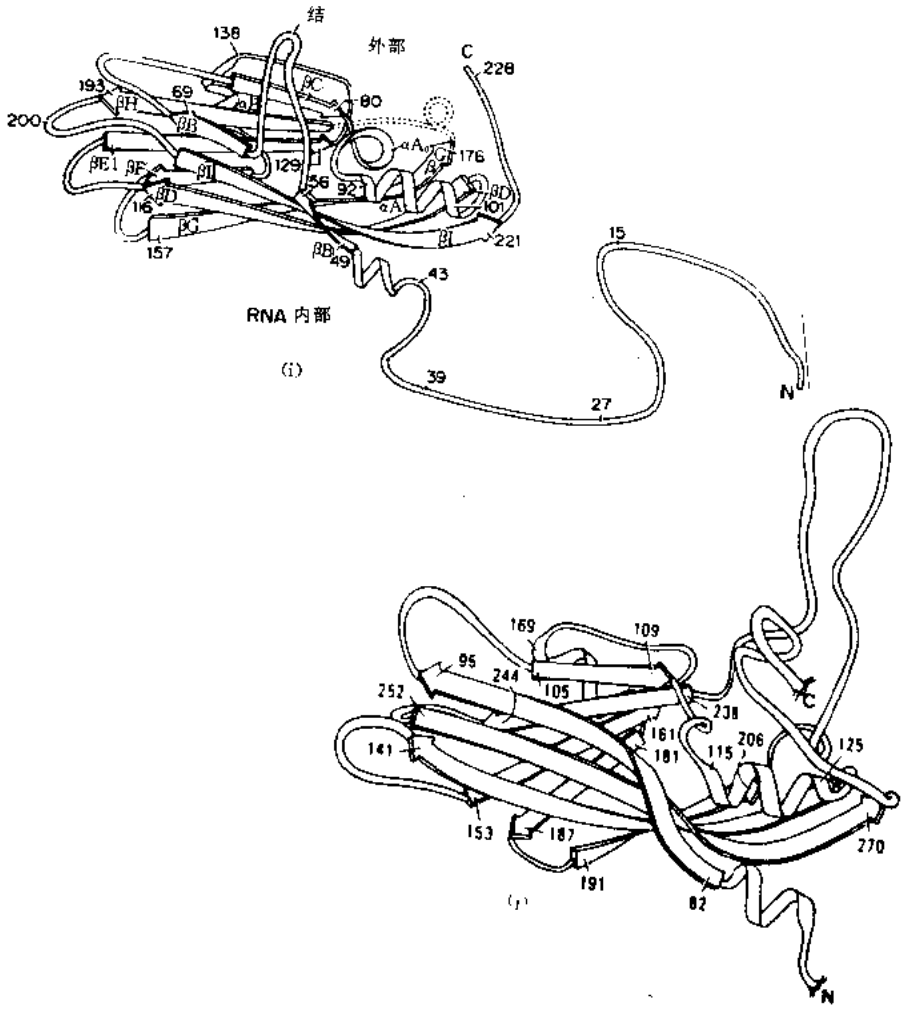
(f)


105
85
II
外部
I
FMDV N-Ag
211
C
βC
161
259
βH
αB
232
βB
βG
βE1
132
βF
βI
αA
120
βD
βD
离子位点
βG
185
βB
1
N
38
RNA 内部
32
20

(g)


161
外部
153
PUFF
C 256
171
220 βH
βC
143
αB
βB
βE
121
βF
βI
131
αA
βD
206
103
49
187
βG
189
33
βA2
54
βB
ser 11
C
20
βA1
VP4 70
14
RNA 内部
N
4

(h)


结
138
外部
C
228
193
βH
69
βC
80
200
αB
βB
αA
176
βE1 βF
βI
129
56 92
βD
101
16
αA
βD
βG
βI
221
157
βB 49
43
RNA 内部
15
39
27
N

(i)

169
109
95
244
105
252
238
161
206
125
141
181
115
153
187
270
191
82
N

(j)

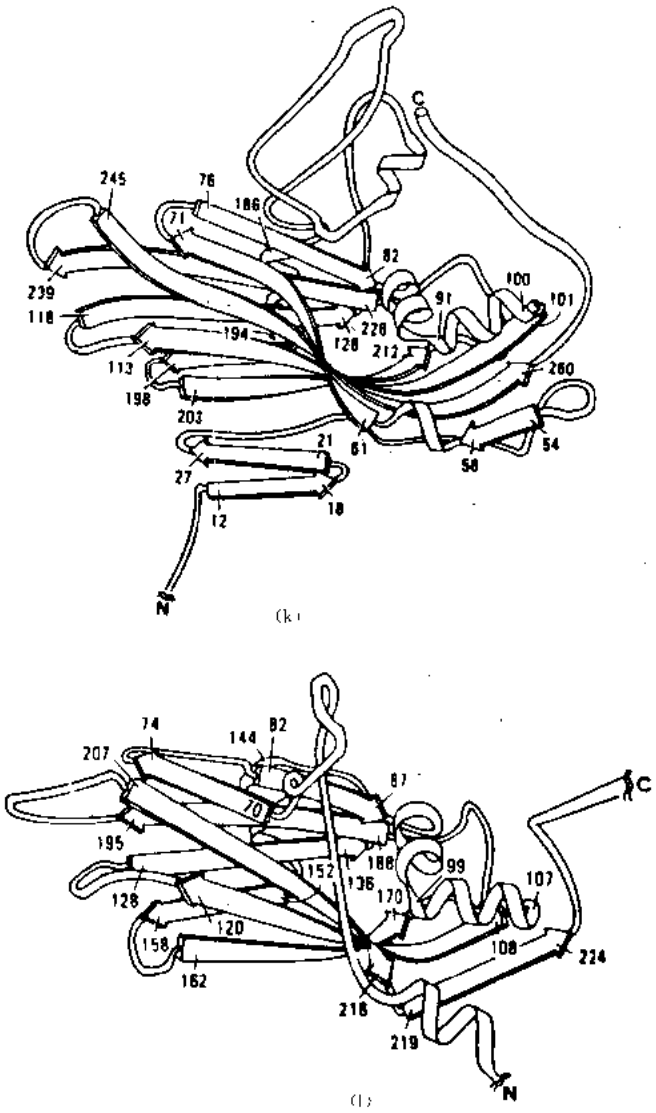


图9.20　各种球形病毒外壳蛋白质的亚基或结构域与 TBSV 的 S 结构域的比较

(a)TBSV 的 S 结构域;(b)SBMV 的 S 结构域;(c)STNV 的 S 结构域;(d)、(e)、(f)分别为 HRV14 的 VP1,VP2,VP3的 S 结构域;(g)、(h)、(i)分别为 Mengo virus 的 VP1、VP2和 VP3;(j)(k)(l)分别为 Poliovirus 的 VP1、VP2和 VP3

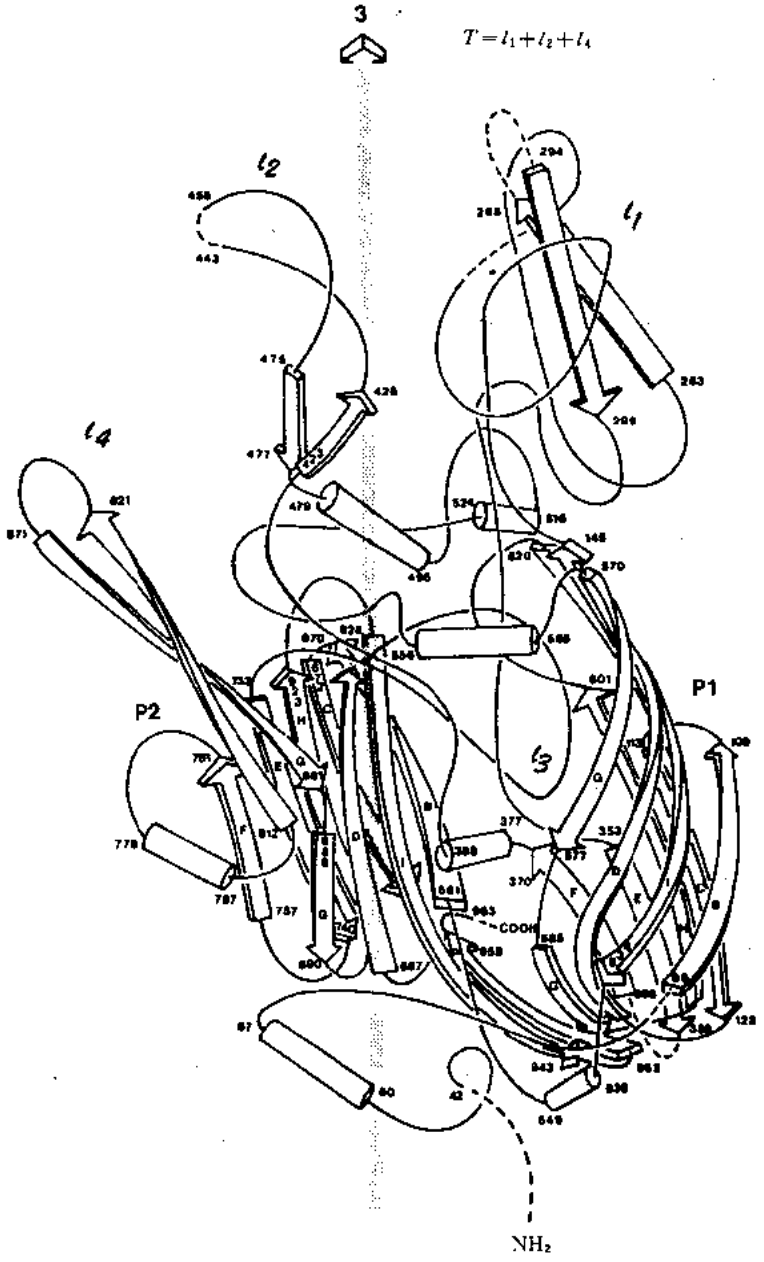


图9.21　腺病毒的六位体亚基及其各个结构域P1、P2和T的结构

图中还标明了六位体的三重旋转轴

第十章　核小体和染色体的立体结构

像化学家和物理学家假设了看不见的原子和电子一样，早年的遗传学家也假设了看不见的遗传要素——基因(gene)，并早已证明，细胞学中用染色法可以看到的染色体(chromosomes)是基因的携带者。现代生物科学技术的发展，进一步证实了所有原核生物和真核生物的遗传基因都是由 DNA 构成的；在病毒中，有些病毒的基因是 DNA(DNA 病毒)，有些病毒的基因是 RNA(RNA 病毒)。

图10.1　染色质核线的电镜照片
其中珠状颗粒的直径约为10nm

我们知道，天然存在的双螺旋 DNA 分子是很长的。原核生物的染色体较为简单，是由裸露的 DNA 双螺旋分子所组成的。例如大肠杆菌的染色体是一个由400万个碱基对连接成环的双螺旋 DNA 单分子，其分子量为2.6×10^9道尔顿。但是，真核生物比原核生物含有更多的遗传信息，例如一个果蝇染色体为5.8×10^{10}道尔顿。而且，与低等生物不同，真核细胞染色体中的 DNA 并非裸露的，而是与一类小的碱性蛋白质即组蛋白紧密结合成核蛋白。绝大多数的真核基因具有染色体型的组织结构，比原核生物的基因组具有更为复杂的结构。

真核生物染色体的结构十分复杂，而且在光学显微镜下只有在细胞有丝分裂的过程中借助于孚尔根染色法才能见到。在电子显微镜照片中可以看到它是由细胞中的一种细丝状物质多重折叠浓缩而成的，这种直径约为30nm 的细丝状物质称为染色质(chromatin)。染色质经过核酸酶简单消化后，在电镜中可观察到它成为直径约10nm 的细纤维，称为核线(nucleofilaments)，见图10.1。在这些核线中有一些直径约10nm 的被称为核小体(nucleosomes)的小球由一细索相连。

生物化学研究证明，几乎所有高等生物细胞染色体都含有 DNA、组蛋白、染色体非组蛋白和少量 RNA。一个真核染色体含有一个双螺旋 DNA 分子。染色质中的组蛋白共有5种，即 H_1，H_{2a}，H_{2b}，H_3和 H_4，都是分子量为11000—21000道尔顿的碱性蛋白质。染色体非组蛋白有100多种，其中有的是酶，有的是酸性蛋白质，但不都是酸性蛋白质。

生物化学研究，尤其是用核酸酶消化染色质的实验表明，染色质是由核小体的重复单位组成的。核小体中有一个长度相当恒定的 DNA 分子片段，常常是200个碱基对。DNA 与组蛋白的八聚体相结合。八聚体中含有 H_{2a}，H_{2b}，H_3和 H_4各一对。另一个组蛋白 H_1也与核小体相联系，但联系的方式尚不清楚。用核酸酶进一步水解核小体后，

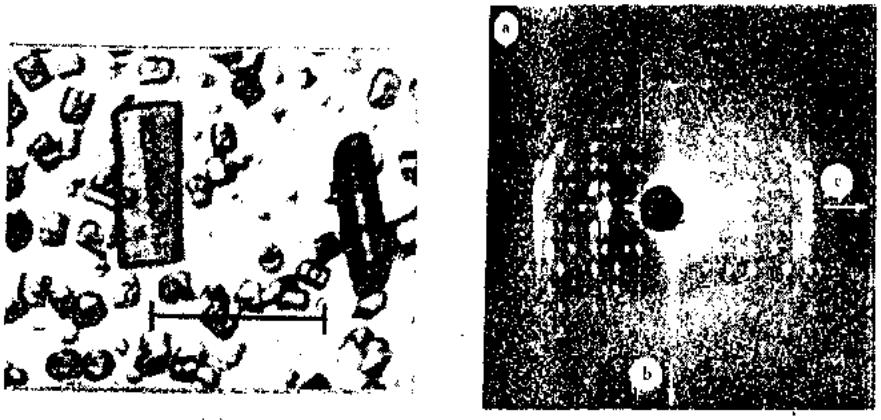
产生了一个核小体"核心"颗粒(nucleosome core),在至今所研究过的细胞类型中,核心颗粒所含的DNA片段长度都是140个碱基对。其余的不包含在核心颗粒中的DNA被认为是连接相邻两个核心颗粒的,因而被称为连接DNA(linker DNA)。

关于染色体的立体结构研究,除了用电子显微镜和中子衍射等方法外,从20世纪50年代至今已用X射线衍射技术对染色体的不同结构层次的样品进行了大量研究工作。从活细胞、分离的细胞核以及细胞分裂中期的染色体的X射线低角衍射[117,118],染色质纤维的X射线纤维衍射[119,120,121],直到核小体核心颗粒的晶体[122,123,124,125]以及核心颗粒中的组蛋白八聚体[126]的晶体的X射线单晶衍射结构分析,成为揭示染色体立体结构的最有效的技术。

10.1 核小体核心颗粒的结晶和单晶结构分析

由于染色体又大又复杂,因而不容易也未得到结晶。核小体核心颗粒则已得到晶体,不过由于晶胞相当大,也是X射线衍射法难以对付的。因此,常常还应用电子显微镜直接观察测量,并用电子显微镜傅里叶三维重构技术以及中子衍射等方法与之相配合进行结构分析。

1977年,J. T. Finch 和 A. Klug 等人报道了他们已获得大鼠肝脏核小体核心颗粒的晶体,



(a) (b)

图10.2 核小体核心颗粒的晶体(a)和旋进照片(b)

并开始了X射线单晶衍射结构分析。小的单晶体及其衍射照片见图10.2。晶体生长条件为:0.5—2mg/ml 核小体核心颗粒、0.4—0.6mmol/l 精胺四氯化物(或3—4mmol/l $MgCl_2$以及0.2mmol/l 精胺;或10mmol/l $MgCl_2$、1.5mmol/l 精胺以及100mmol/lNaCl)在10%(W/V)1,6-己硫醇(作为沉淀剂)溶液中,在4℃或室温下经过1—4天,即能得到小的单晶或微晶。可以用坐滴式汽相扩散法生长晶体,也可以将适当的上述混合液装入玻璃管中,然后将管封塞,晶体在管中生长。晶体的X射线衍射旋进照片见图10.2。分辨率只能达到2.0nm左右。收集了2.5 nm分辨率的衍射数据。晶体属正交晶系,空间群为$P2_12_12_1$,$a=11$nm,$b=19.2$nm,$c=34$nm,$\alpha=\beta=\gamma=90°$。一个不对称单位中含有三个分子量约为200000道尔顿的核心颗粒。由于晶胞很大,仅用X射线衍射法难以解出其结构。用电子显微镜三维重构技术作为辅助方法,将晶体碎片用负染法拍摄电镜照片,它相当于X射线衍射法中实空间电子密度投影图。因而可用光学

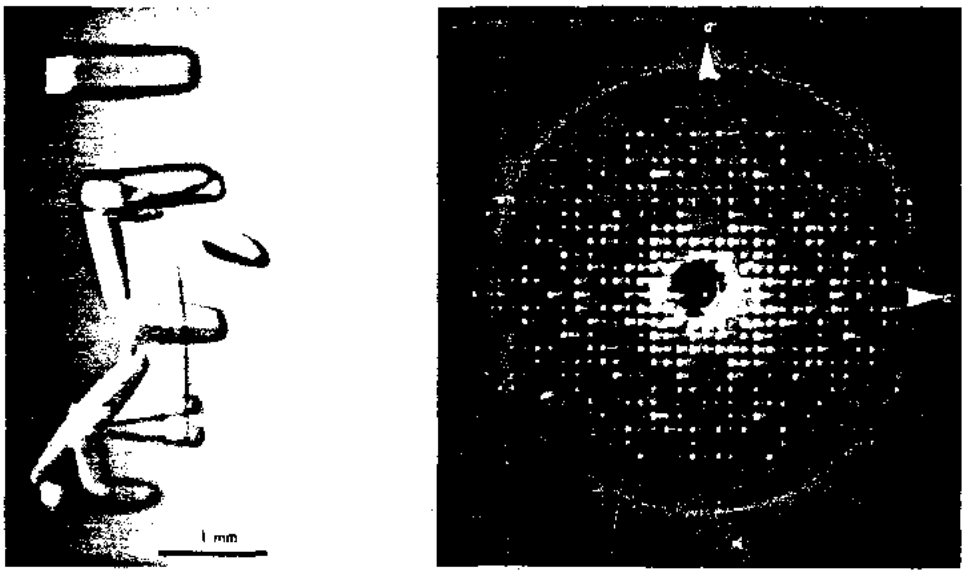
衍射和计算机傅里叶变换法计算出相应于X射线衍射法中倒易空间的中心截面投影图(包括各个衍射点的振幅和相角符号)。将由电镜术确定的相角符号与X射线衍射法收集到的衍射点的强度(计算出结构因子的振幅)数据结合起来,计算出晶体沿 a,b,c 三个方向的电子密度投影图及三维帕特逊图,解释了核心颗粒在晶胞中的堆积。图10.3(a)为晶体在晶胞 a 方向的电子密度投影图,在此方向投影图中核心颗粒相互重叠最少,分辨率为2.5nm。后来发现,在这种核小体核心颗粒的晶体中,核心颗粒的组蛋白已部分地被酶水解了。不过,它们的物理化学性质仍然保持与完整核心颗粒的性质十分相似。

图10.3 核小体核心颗粒晶体沿 a 方向的电子密度投影图

(a)大鼠肝部分被水解了的核小体核心颗粒晶体的电子密度投影图,晶胞较大,沿 c 轴有6层;

(b)牛肾完整的核小体核心颗粒晶体的电子密度投影图,晶胞较小,沿 c 轴有2层

1981年J. T. Finch 和 A. Klug 等人报道了从不同种属的染色质(大鼠肝、鸡红血球、牛肾、小牛胸腺、扇贝精子以及小鼠骨髓瘤细胞)得到了完整的核小体核心颗粒的新晶型。这些晶体和它们的X射线衍射花样十分相似,表明了核小体核心颗粒的普遍特征。图10.4中为牛肾的完整核小体核心颗粒晶体及其旋进照片。这些晶体的晶胞较前述被酶部分水解了的核心颗粒晶体的小,有序性高得多。在 c^*,a^* 和 b^* 方向旋进照片上分辨率分别可达0.5nm、0.6nm 和0.8nm。最好的晶体的生长条件与前述的相似,但是不加有机溶剂1,6-己硫醇(虽然在安装晶体进行衍射实验前,加入5%(W/V)1,6-己硫醇以增加稳定性)。这些新晶型的空间群仍为 $P2_12_12_1$,晶胞参数为:a=11.1nm,c=11.1nm,b=19.3—20.3nm。每个不对称单位中含有一个分子量为206000道尔顿的核心颗粒。晶胞的大小依赖于在特殊的制备样品方法中核小体核心颗粒的DNA长度以及晶体的水化程度。发现当收缩到一个基本上稳定的最佳状态时,晶体的有序性增加。例如调节1,6-己硫醇的浓度为21—27%时,得到了晶胞参数为 a=10.75nm,b=18.4nm,c=11.05nm 的晶体。用低角度较强的X射线衍射照片0kl 推导出来的相角符号及X射线衍射强度计算出电子密度投影图,见图10.3(b)。1984年 T. J. Richmond, J. T. Finch 以及 A. Klug 等人报道了用牛肾核小体核心颗粒的这种新晶体所得到的0.7nm 分辨率晶体结构。



(a)　　(b)

图10.4　牛肾的完整核小体核心颗粒的晶体(a)及其旋进照片(b)

将晶体装入含有低成胶温度的琼脂糖的溶液中，在4℃下收集衍射数据，以保持恒定的水化度并固定位置。收集到3.5—0.7nm 分辨率共3300条反射及它们的福里得反射对。用同晶置换法测定相角。为了克服这么大的不对称单位以及衍射强度较小所产生的困难，重原子衍生物的制备选用了一些含有若干个重原子的簇化合物。最后用了两个重原子衍生物，一个是 TAMM [即 tetrakis(acetoxymercuri)methane]，另一个是 PIP[即 di-μ-iodo,bis(diaminoethyl-N,N′) diplatinum(Ⅱ)]。计算出的电子密度图已可清楚地看到 DNA 超螺旋的双螺旋等。

10.2　核小体核心颗粒的立体结构

从 Finch 和 Klug 早年的低分辨率电子密度图中就可清楚地看到核小体核心颗粒是一个 11.0nm×11.0nm×5.7nm 的扁圆体，略为楔形。并可假定，组蛋白八聚体在中心，DNA 是右手 B 型双螺旋，它再绕着蛋白质八聚体盘绕1.75圈，形成左手盘绕的超螺旋，螺距为2.8nm。现在已被公认的核小体的三维结构模型即是在此基础上建立的，见图10.5。当然在分辨率仅为 2.5nm 的分析阶段，是不可能跟踪出 DNA 在核小体中的走向的，也不能肯定地判断组蛋白的排列。他们提出组蛋白可能排列成两层或两圈。核小体核心颗粒中的4种组蛋白 H_{2a}，H_{2b}，H_3和 H_4各自成对地存在，已提示了核心颗粒可能具有二重轴对称性。

A. Klug 等人对核心颗粒晶体进行的 X 射线单晶结构分析以及他们对核心颗粒中的组蛋白八聚体的有序聚合物负染样品进行的2.2nm 分辨率电子显微镜像的三维重构研究，以及中子衍射和化学交联等实验结果，证实了组蛋白八聚体确实具有一个二重轴对称性。八聚体的整体形状像一个左手型的线轴，在核小体中约有两圈 DNA 超螺旋缠绕于其上。$(H_3\text{-}H_4)_2$四聚体形成了一个盘，这个盘结合了并限定了 DNA 超螺旋的中央圈；而两个 $H_{2a}\text{-}H_{2b}$二聚体则位于这

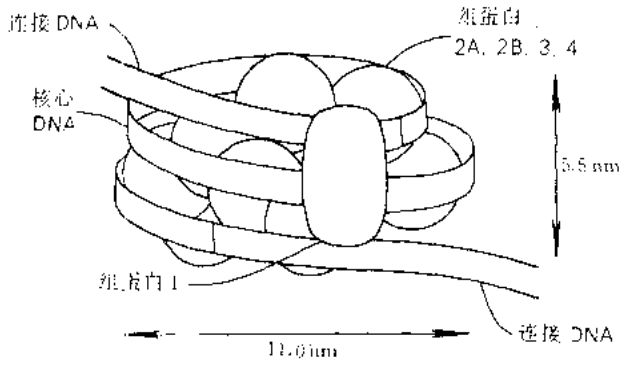


图10.5　核小体的结构模型

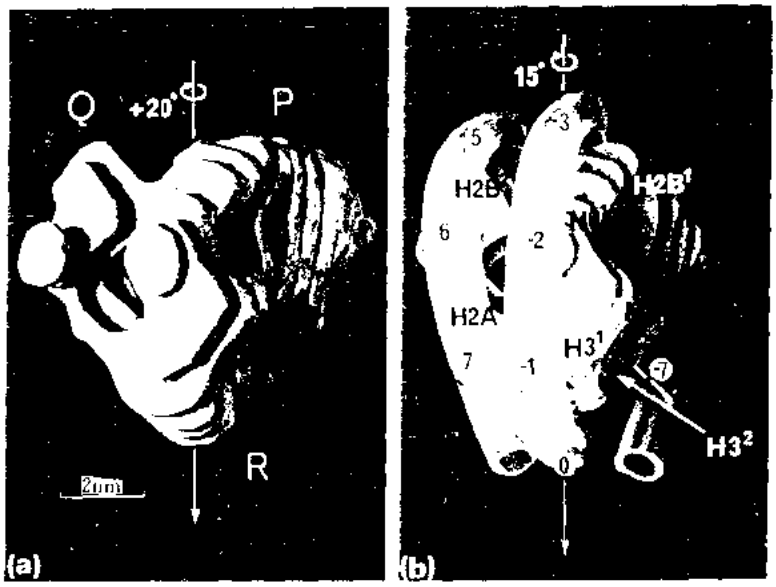


图10.6　组蛋白八聚体的模型(a)和有两圈 DNA 超螺旋缠绕于其上的组蛋白八聚体模型(b)
箭头表示二重轴。沿着 DNA 的距离用数字－7至＋7表示(以二重轴通过处作为原点0)以标出核心颗粒中 DNA 的146个碱基对中所含有的双螺旋的14个重复。组蛋白分子名称中的上标是按照 DNA 位置的顺序给定的

个盘的两侧的两个面上，并各与大约半圈 DNA 超螺旋相联系，见图10.6中的模型。T. J. Richmond，J. T. Finch 和 A. Klug 等人关于完整核心颗粒的晶体的0.7nm 分辨率 X 射线衍射晶体结构分析，证实了前述的模型，并进一步使人们真正见到了结构的许多细节，揭示了许多新的结构特点。

10.2.1　DNA 双螺旋和超螺旋

在0.7nm 分辨率电子密度图中可清楚地看到 DNA 的深沟和浅沟(磷酸酯骨架)交替出现，见图10.7。表明了核小体核心颗粒中的 DNA 是右手 B 型双螺旋，并进一步折叠成一个左手的超螺旋。可以清楚地看到12圈 DNA 双螺旋，仅在两个末端处大约有半圈不能辨认，见图10.8中的模型。

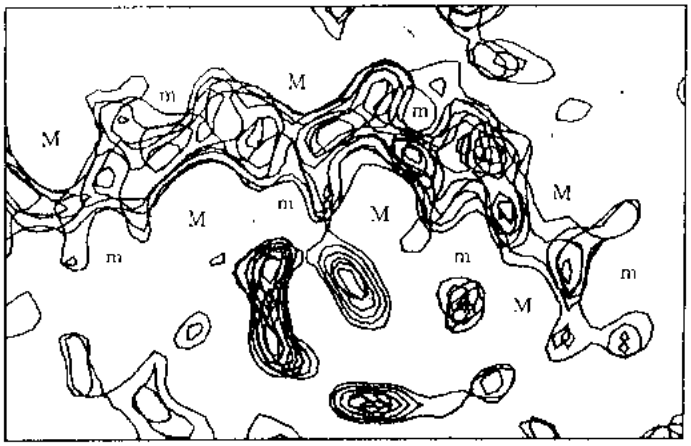


图10.7　DNA 超螺旋的一段弧形电子密度

此图是用通过这个双螺旋的3个截面(−10、−11、−12,间距为0.173nm)堆积而成的,图中可看到双螺旋的两圈,标出了交替的深沟(M)和浅沟(m)。在其下方的电子密度是 H_{2B} 的部分

在一整圈超螺旋中,平均大约有7.6圈双螺旋,相邻两圈超螺旋以一个浅沟几乎对着一个深沟的方式排列着。由中央圈的位置−3.8到+3.8测定出的超螺旋的螺距约为3.0nm;由位置−2.5到+5.0测出的值为最小,约2.5nm;而若沿着超螺旋的一段弧所测的螺距则大约为3.5nm。在0.7nm 分辨率下,尚不能测定每圈双螺旋中实际的核苷酸数目,估计达到0.5nm 分辨率时应该能看到 DNA 磷酸盐骨架扭曲的精确性质。因而才可以与标准的 Watson-Crick 双螺旋进行比较。但是,由图10.8可以看到,DNA 双螺旋绕着组蛋白八聚体盘旋时并不是形成一个非常规则的超螺旋,而是在若干个位置处急剧地弯曲或者可能缠结。在位置+1的区域中,DNA 的弯曲是自身的弯曲而不是在与组蛋白直接接触的方向上,但是是由于在位置+0.5处与组蛋白的接触所致;在位置+4处也具有相似的弯曲,而且明显地也是由于在位置+3.5处的 DNA-组蛋白接触。在二重轴对称性联系起来的点−1和−4处也有相似的 DNA 弯曲。从图10.8中还可看到,深沟和浅沟的宽度似乎依赖于弯曲的程度。测量结果表明,在+1处,浅沟的宽度在核心颗粒内面为0.7nm,在外面增加至1.3nm;与此相似,深沟的内外宽度分别为1.1nm 和2.0nm;在最弯曲的区域中,颗粒内面上深沟比浅沟更为紧缩。最紧凑的 DNA 弯曲位置±1和±4正好是不被 DNA 酶 I 消化的最佳位置。由于 DNA 外侧没有蛋白质密度,所以 DNA 不同部位核酸酶酶解速率不同的问题不能用简单的空间障碍效应加以解释,而很有可能是由于 DNA 因形成超螺旋所产生的构象变化,或是由于 DNA 与组蛋白紧密结合而降低了柔性。

10.2.2　组蛋白八聚体及其与 DNA 的相互作用

0.7nm 分辨率电子密度图中,在 DNA 超螺旋的内侧可以清楚地看到若干个高电子密度的棒状区域,很可能是蛋白质的 α 螺旋,显然是组蛋白的位置,见图10.7。组蛋白的电子密度看来没有包围到 DNA 的外侧,也没有插入到 DNA 超螺旋的各圈之间。组蛋白与 DNA 超螺旋的一端到另一端的主要接触顺序是:H_{2a},H_{2b},H_4,H_3,H_3,H_4,H_{2b},H_{2a},并呈二重轴对称关系,与图10.6的低分辨率模型是相符的。

组蛋白几乎与每一圈 DNA 双螺旋都有接触,而且这些接触都仅限于在 DNA 超螺旋的内表面。其中,如图10.9所示,组蛋白 H_3 与 DNA 超螺旋中央圈位置−3.5和+3.5处的浅沟有明

显的接触。此外，组蛋白 H_4 也与 DNA 位置从±4到±5有次要的接触。H_{2a}·H_{2b}二体与 DNA 位置－3.5、－4.5以及－6.5处的浅沟形成接触。

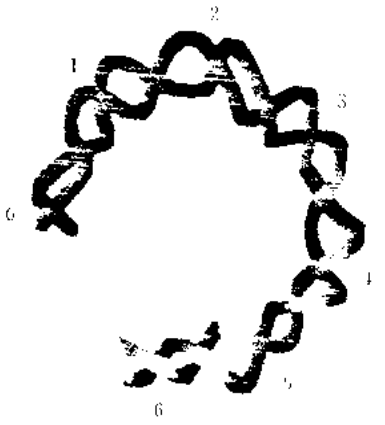


图10.8　用磷酸酯骨架所表示的 DNA 双螺旋折叠成超螺旋的模型
为了清楚起见，只画了沿超螺旋的轴方向观看到的一圈超螺旋。分子的二重轴为通过位置0处的浅沟的水平方向的轴

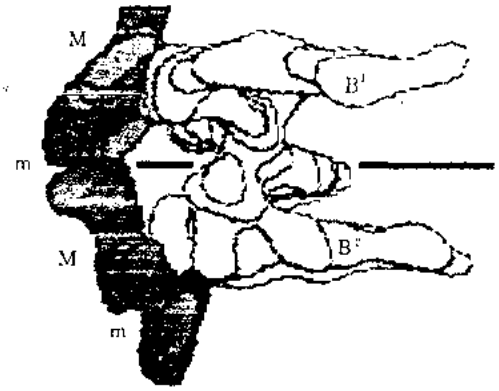


图10.9　H_3二体及其与 DNA 超螺旋的中部区域的接触
黑色处代表 DNA，白色处代表 H_3

10.2.3　连接 DNA 与组蛋白 H_1的结合

我们知道，在染色质中相邻核小体之间是由连接 DNA(linker DNA)连接的，它的结构尚不清楚，而且可能是变化的。在核小体核心颗粒晶体的结构分析中也不可能包含它的结构，但是在电子密度图中观察到 DNA 位置0的两侧空间被 H_{2a}分子所填充了。如果溶液中也是这样的话，那么连接 DNA 将会伸延到位置＋7和－7以外，因而不可能是简单地沿着由 DNA 中央圈所形成的平滑的超螺旋的路径。此外，由于 H_{2a}分子伸展到颗粒的外部边缘，因而会使连接 DNA 与分子的二重轴形成大约50—60°的角度。而且由于染色质含有的连接 DNA 都很短(即0—10个碱基对)，因此很有可能要求 DNA 急剧地弯曲或缠结。从 DNA 在核心颗粒末端的排列可推测 H_1可能与超螺旋的两个末端相结合，从而形成超螺旋到连接 DNA 的过渡。

10.3　染色质的 X 射线纤维衍射分析和染色质超级结构的螺线管模型

我们知道，在细胞分裂中期所看到的染色体是由染色质高度浓缩而成的。也就是说，染色质具有更高的超级结构，并且早已成为 X 射线衍射方法研究的课题。用这一方法已经对各种不同浓度的染色质胶、各种 pH 值或金属阳离子的影响以及保留或除去组蛋白 H_1等情况下产生的衍射花样作了大量的研究和分析。在此，只简单介绍最重要的结果和公认的模型。

10.3.1　染色质胶纤维样品的制备和 X 射线衍射方法

鼠肝细胞核经微球菌核酸酶简单消化后所制备的天然染色质，长度约为10—100个核小

体，平均为40个。A. Klug 等人用这种样品配制在0.2mmol/l $MgCl_2$、5mmol/l Tris-HCl，pH 7.4，0.1mmol/l NaN_3，0.2mol/l 苯甲磺酰氟溶液中。30000r/min 离心2—12小时，得到染色质浓度约为20%(W/W)的离心沉降物。然后立即用一个细长的玻璃棒将这种染色质胶推入一个1.0mm 的硅化了的玻璃毛细管中，并用腊封上毛细管。在毛细管上涂一点 BaF_2，它将给出尖锐的0.358nm反射，作为测量的标准参照。用 X 射线纤维衍射法拍摄衍射照片，见图10.10。样品与底片的距离为11或15cm，曝光2—4天。

图10.10 20%(W/W)的染色质胶的 X 射线纤维衍射花样

10.3.2 染色质的螺线管模型

在衍射花样中，可以看到有一个明显的尖锐的11.0nm 反射，此外还有5.7nm，3.8nm 和2.8nm 的反射等。虽然只有很少几条反射，但是根据生物化学知识和电镜观察中看到的螺线管结构等资料，A. Klug 等人证明了上述 X 射线衍射花样中的11.0nm 反射是由螺线管的各圈之间的重复的间距

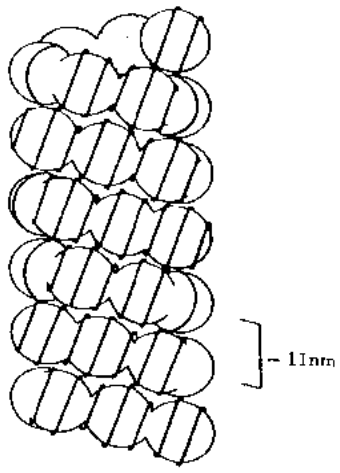


图10.11 染色质的螺线管模型
核线折叠成螺线管形，螺距约为11nm。每个椭圆体代表一个核小体，外部盘绕的粗线代表 DNA 双螺旋

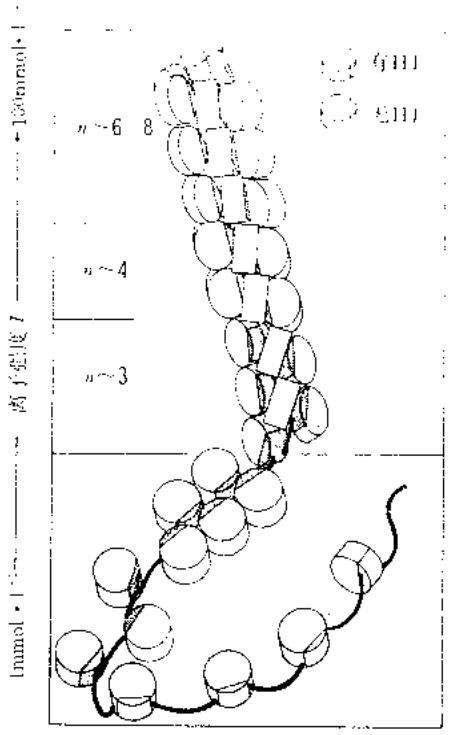


图10.12 在有组蛋白 H_1存在和无 H_1的情况下染色质超级结构的示意图

所产生的，而不是由于核小体(直径约10nm)之间沿着核线的重复间距产生的。也就是说，螺线管各圈之间的间距为大约11nm，因而为染色质的超级结构提出了一个螺线管模型，见图10.11。此外，实验还表明，这种螺线管结构需要组蛋白 H_1存在，才能稳定。如果 H_1不存在，则不能形成高度有序的螺线管结构，见图10.12中的示意图。图中 n 表示螺线管每圈的核小体数

目。图的左下方表示在低离子强度下含有 H_1 的染色质的超结构是相当不规则的，随着离子强度增加，逐渐形成规则的螺线管结构，而每圈中所包含的核小体数目也随之增加。图的右下方表示不存在有 H_1 时，则观察不到高度有序的结构。

染色质的这种螺线管结构模型与在电镜下观察到的核中的30nm 直径的“线”是有关系的。这些螺线管是否再进一步浓缩以及染色体中其他组分如非组蛋白等起什么作用等问题，待进一步研究。

第十一章　多糖的立体结构

在生物的组成物质中，糖类是最早被人们所研究和认识的。单糖、双糖、寡聚糖等小分子以及淀粉、纤维素等多聚物大分子，都是天然物质。天然糖类往往是动植物结构材料的主要成分以及营养特别是能量贮备的主要来源。19世纪末和20世纪初，已经弄清楚了它们的化学组成和结构，并早已用X射线衍射方法来研究它们的立体结构。其中作为天然大分子的淀粉和纤维素，由于都是由一种单糖(葡萄糖)缩合而成的，结构相对地来说较为简单，在20世纪30年代就开始用X射线衍射技术初步研究了它们的立体结构(虽然现在看来许多模型不完全正确)。后来人们以为关于糖类的问题，包括结构和生物功能等，都已经基本上解决了，因而糖类的研究工作在相当长的一段时间内不为人们所重视，研究工作未得到深入的发展。但是，在1970年前后，由于免疫学研究和生物膜研究的进展，例如指出了抗体是一种含有共价连接的糖类的蛋白质大分子；了解了生物膜上存在着若干寡聚糖链，糖蛋白和糖脂在生物膜中具有重要地位等。而且由于越来越多的复合多糖被发现，例如血浆蛋白质和许多酶都是糖蛋白，神经节则有种类繁多的糖脂，以及很多具有重要生物功能的物质如核酸、神经节苷脂、某些抗菌素、毒素、凝集素以及血型活性物质等，均属于含糖的复合物等等，推动了人们重新把兴趣转移到糖类的研究上。研究的起点，仍然是关于糖类的化学结构和立体结构。

实际上，与蛋白质和核酸相比，多糖的结构测定包括化学结构和立体结构的分析看来是最难解决的。对于X射线衍射方法来说，这首先是由于制备合适的样品是困难的；而且，由于得不到很好的衍射图，因而需要用大量的计算机机时来评价各种可能的模型的正确性。这也就是为什么虽然人们最早是从糖类的结构开始研究生物的组成，而且从20世纪30年代就对多糖进行了X射线衍射研究，但至今尚未有确切的结构模型的原因。

多糖是由一种或几种单糖或单糖的衍生物以糖苷键相连接成的生物大分子。最常见的是以D-葡萄糖为建造单元的多糖，如纤维素、淀粉、糖元等，见图11.1。此外，作为多糖的建造单元的其他单糖或衍生物还有如D-甘露糖、D-或L-半乳糖、D-木糖、L-阿拉伯糖；D-氨基葡萄糖、D-氨基半乳糖、D-甘露糖胺丙酮酸以及D-葡萄糖酸、D-半乳糖醛酸和D-甘露糖醛酸等。

11.1　多糖的构象

与蛋白质的多肽链和核酸的糖-磷酸骨架相似，多糖链中也有几个可以自由旋转的构象角，因而也可具有各种各样的构象，形成稳定的三维立体结构。在本书中，我们主要以纤维素和淀粉的结构为例加以说明。

11.1.1　吡喃糖的折叠构象

六碳糖的环形结构为吡喃环，它可以两种构象即所谓的椅式和船式存在，见图11.2。通常椅式比船式稳定得多。环碳上的10个取代的原子团按空间位置可分为两类，一类几乎垂直于环平面，称为直立键(axial bond)，一般用虚线表示；另一类几乎平行于环平面，称为平伏键(equatorial bond)，用实线表示。在平伏键位置处的取代为能量低的状态，因为与其他取代基

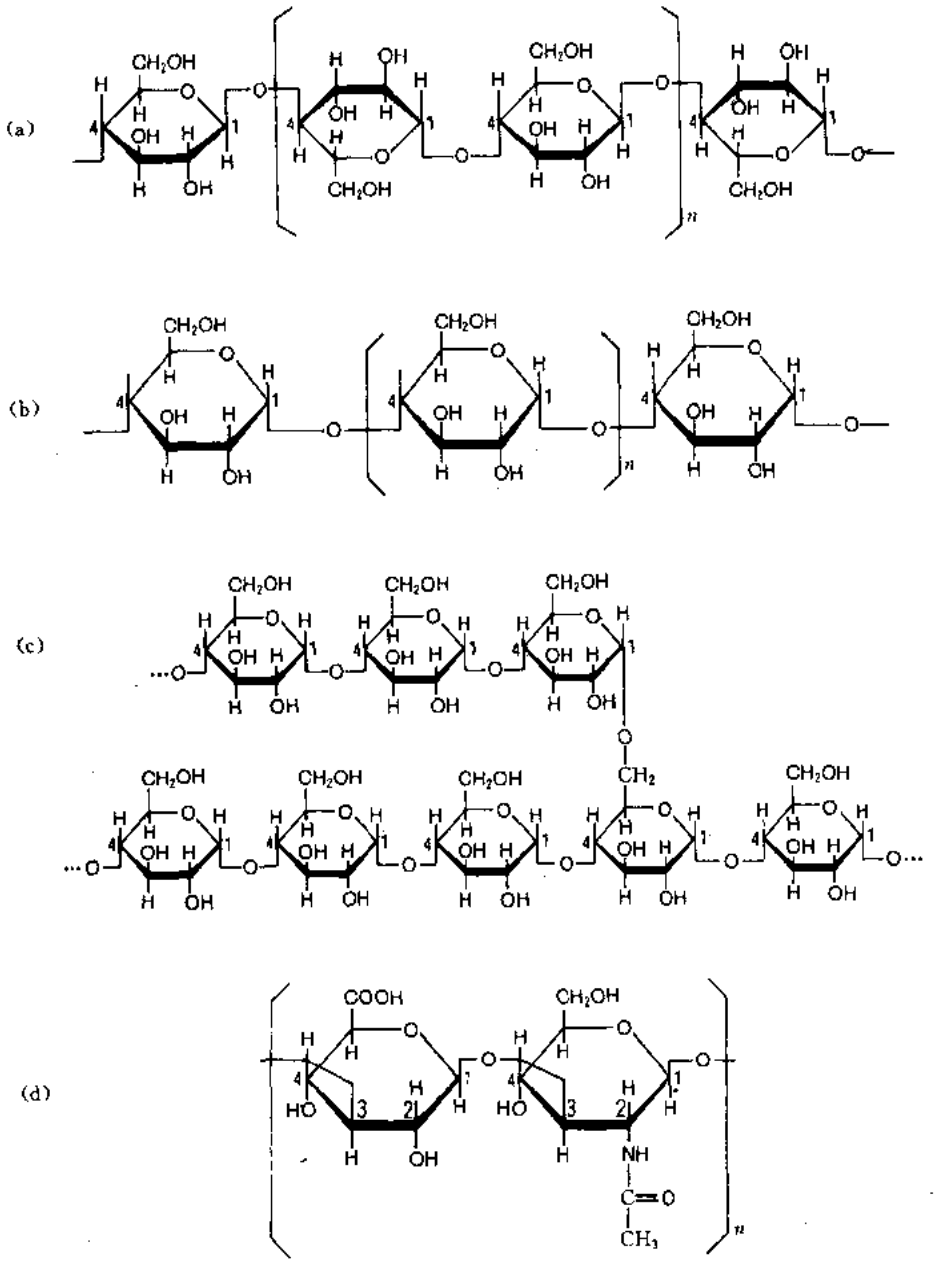


图11.1　几种常见多糖的化学结构

(a)纤维素(β-1,4-糖苷酯键);(b)直链淀粉(α-1,4-糖苷酯键);(c)支链淀粉(α-1,4-糖苷酯键以及 α-1,6-糖苷酯键);(d)透明质酸(由 D-葡萄糖醛酸和乙酰-氨基-葡萄糖以 β-1,3-糖苷酯键相连接)

团之间较少有立体障碍,这对于大的取代基团来说更为重要。图11.3表明了两种异头物(anomere)即 α-D-葡萄糖和 β-D-葡萄糖的椅式构象,还可看到比 H 大的—OH 基团和—CH_2OH基团都是在平伏键的位置。

11.1.2　多糖链的构象角

多糖中的单糖单元以刚性的环式存在,如 D-葡萄糖以椅式的吡喃环型式存在。正如可将刚性的肽键平面基团看成多肽链中的结构单元那样,按照 IUPAC-IUB(国际基本与应用化学

图11.2　吡喃糖的椅式(a)和船式(b)折叠构象

图中 a 表示直立键，e 表示平伏键

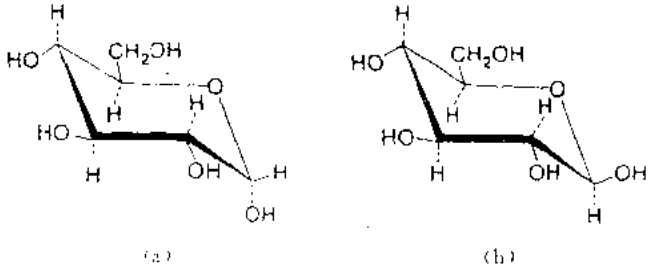


图11.3　D-葡萄糖的两种异头物

(a)α-D-葡萄糖；(b)β-D-葡萄糖

联合会-国际生物化学联合会)的定义，在多糖链中，以1,4-连接的糖苷键，有两个可以自由旋转的扭角 ϕ 和 ψ；以1,6-连接的糖苷键，有三个扭角 ϕ、ψ 和 ω，见图11.4。从图中可以看到，由于糖苷酯键的类型和糖环上取代基团等不同，多糖的构象也会不同。例如 β-1,4-糖苷酯键连接的 β-D-葡萄糖多聚物(如纤维素等)可以形成直链，每个葡萄糖单元相对于前一个翻转了180°；而 α-1,4-糖苷酯键连接的 α-D-葡萄糖多聚物(如直链淀粉和糖元等)，由于空间障碍，多糖链自然地转弯，不可能形成直链。

11.2　多糖的 X 射线衍射分析方法

人们已对多种多糖进行了 X 射线衍射结构分析工作，原子坐标已存入蛋白质晶体学库中的已有琼脂糖、软骨素-4-硫酸、软骨素-4-硫酸钙盐、透明质酸钠盐、透明质酸钙盐、角叉胶、硫酸角质素以及被膜多糖等。在此，我们仅以最常见的多糖即纤维素和淀粉为例加以说明。

11.2.1　纤维素的 X 射线衍射分析方法

纤维素是自然界中存在最为广泛的多糖之一，棉花、麻和树木等都含有大量纤维素，例如棉花和麻中90%的成分是纤维素。天然纤维素不溶于水，分子量为5万道尔顿以上，成纤维状。因此，可用 X 射线纤维衍射方法进行结构研究。C. Trogus 等人[127]早在30年代就开始了这一研究。由于纤维素分子的排列具有较好的规则性，因而给出了反射数目虽然不多但却具有特征的纤维衍射图，见图11.5，成为鉴别纤维素的最直接的方法之一。由纤维衍射花样还可以测定晶胞参数和确定空间群等。但由于难以精确测定弧形反射的中心位置，所以晶胞参数有误差。

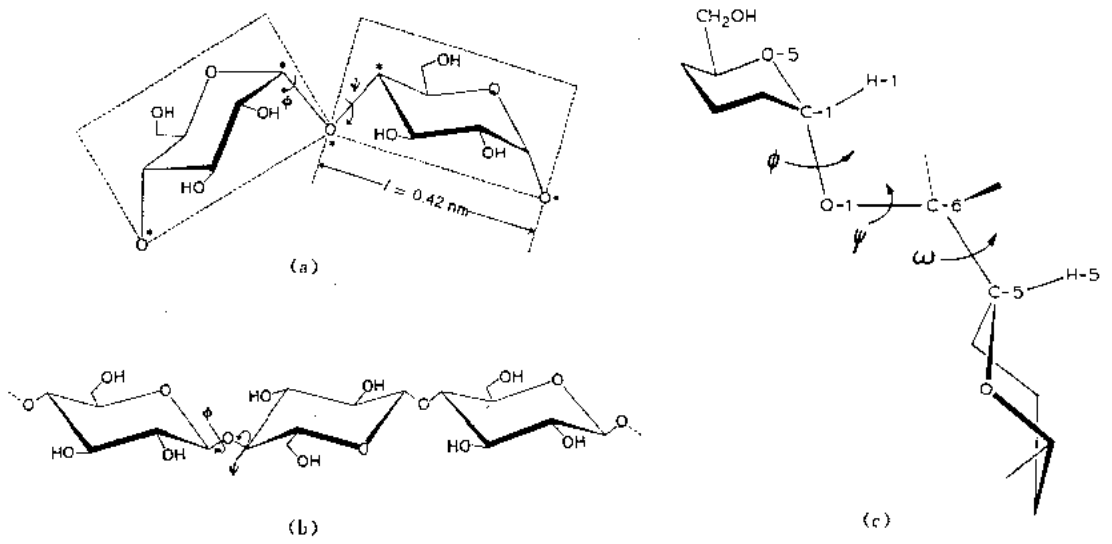


图11.4　多糖链的构象角

(a)、(b)、(c) 分别为以 α 1,4-,β-1,4-以及 α-1,6-糖苷键相连接的 D-葡萄糖聚合体

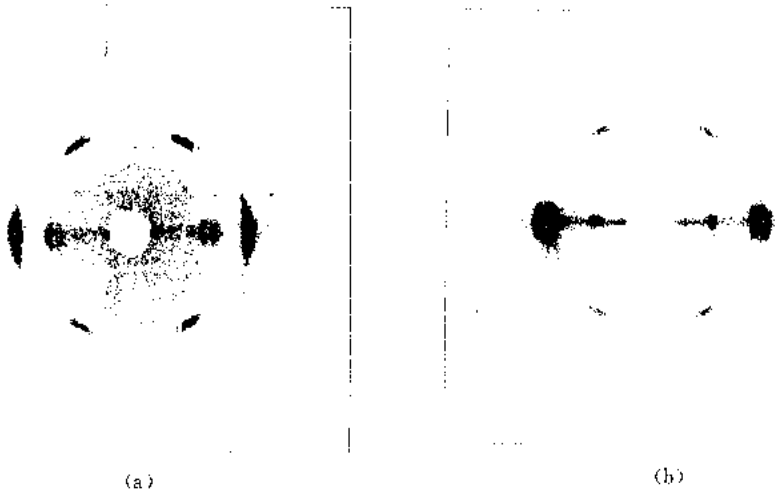


图11.5　天然的(a)和经丝光处理的(b)麻纤维素的纤维衍射照片

样品与底片的距离为51.5mm

根据纤维衍射花样和已知的立体化学及构象知识，可计算和推测纤维素的三维立体结构模型。实验表明，各种来源的天然纤维素的结构是相同的，只是结晶性的程度有所不同。天然纤维素都是链状的，在纤维轴方向的重复距离为1.03nm 左右。关于纤维素分子的立体结构以及它们在晶胞中的排列情况，如究竟是平行排列还是反平行排列等，提出多种模型，然后进行理论计算，并采用最小二乘修正法进行结构参数修正。可将理论值与观察到的衍射花样和衍射强度等相吻合的模型认为是接近真实的纤维素结构和它们的排列方式。

天然纤维素又叫做纤维素Ⅰ，以区别于经过人工处理的各种纤维素和纤维素的衍生物。对纤维素的某些衍生物如硝酸纤维素、醋酸纤维素等以及经过丝光处理的碱的加成化合物即纤维素Ⅱ和经过在氨水中膨胀后的纤维素Ⅲ等也进行了 X 射线纤维衍射研究。其中有些样品可

以给出比天然纤维素更好的纤维衍射花样。

11.2.2 淀粉的样品制备和X射线衍射分析方法

淀粉在自然界中通常以小的颗粒状态存在，因此早期的X射线衍射研究是用天然淀粉粒或用淀粉溶液在不同温度下慢慢蒸发后所制备的样品，用粉末照相法进行的。1930年J. R. Katz等人[128]首次报道了淀粉的X射线衍射花样，发现按照粉末图花样的不同，天然淀粉至少可分为三类，分别称为A型淀粉（谷物的淀粉属于此类）、B型淀粉（块茎和马铃薯等的淀粉属于此类）以及C型淀粉（植物中极为少见的淀粉），见图11.6。后来，又用从淀粉粒中提取并沉淀分离出来的直链淀粉及其各种衍生物的样品或微晶进行粉末衍射照相法研究，得到了一些较好的粉末衍射花样，而且表明了直链淀粉的衍射花样与相应的淀粉粒的相同。粉末照相法是用CuK_α（Ni滤波），样品至圆筒形底片的距离为3—5cm。样品厚度约1mm，放于薄壁的玻璃小室中，曝光照射，收集X射线衍射数据。

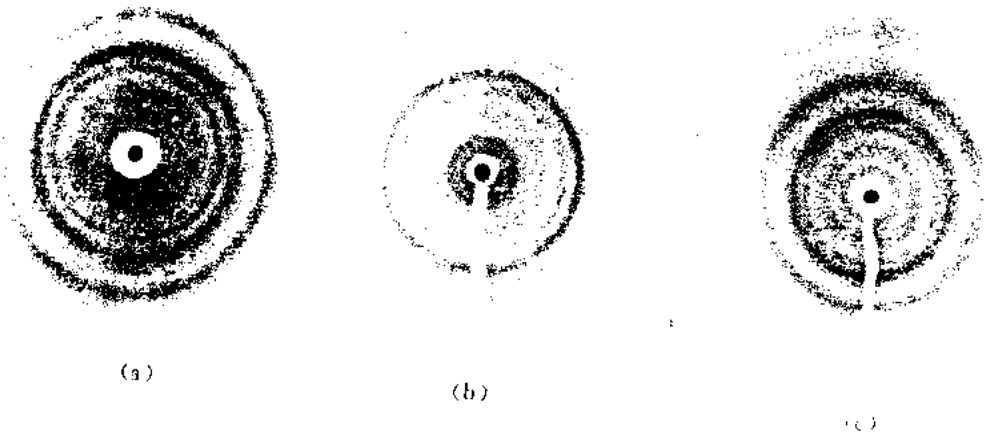


图11.6 A型淀粉(a)、B型淀粉(b)以及C型淀粉(c)的X射线粉末衍射花样

除了用粉末衍射照相法研究淀粉的结构以外，用纤维衍射法可期望能得到比粉末图记录的反射数目更多的纤维衍射花样，因而给出较多的结构信息和较好的结构分析结果。但是，纤维衍射法要求样品中分子是定向排列的。淀粉粒和直链淀粉本身不能给出好的纤维衍射图，而是将直链淀粉先制成非亲水性的衍生物，并将其拉成细丝，然后在保持分子定向的条件下，再使衍生物变回到原来的直链淀粉，即可用纤维衍射法拍摄到直链淀粉本身的X射线纤维衍射图。例如，H-C. H. Wu等人[129]先将直链淀粉进行乙酰化，制成三乙酸纤维素，在170℃下在甘油中被拉长500%，成为分子定向排列的固态细丝。将3—10cm长的细丝绷紧固定在不锈钢钳子上，浸没在0.2mol/l KOH的75%乙醇溶液中24小时，以去乙酰化。所产生的碱直链淀粉在85℃下放置于相对湿度为80%的空气中12天，然后在90℃下于相对湿度为100%的空气中再放置6天，即可形成A型直链淀粉纤维（对于B型直链淀粉纤维来说，则是在与上述同样高的相对湿度中，但放置在常温下，然后在热水中加热而形成的）。用上述纤维在80%相对湿度的空气中拍摄X射线纤维衍射图，见图11.7。为了记录到较高分辨率的反射，可将纤维进行回摆。从纤维衍射图可以计算晶胞参数、测定空间群，并利用立体化学知识提出模型。进行理论计算、能量优化计算等，并与实际观察到的衍射花样和衍射强度进行比较，根据吻合程度判断模型正确

与否。

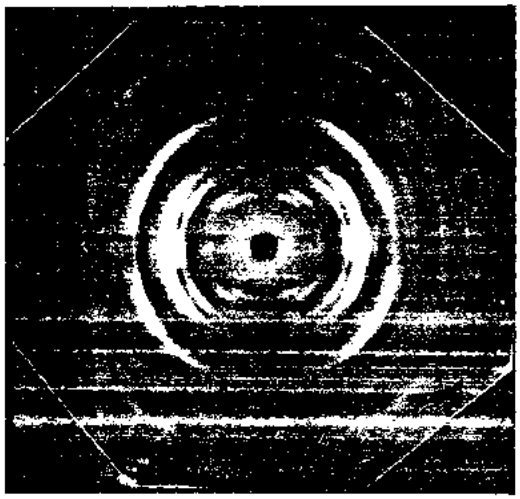


图11.7 A型直链淀粉的纤维衍射照片

11.3 多糖的立体结构

多糖链的构象是多种多样的。现在已经知道的就有几类，如由链形分子所形成的片层结构、由螺旋形分子所形成的空心螺旋结构以及松散结合的链状结构等。我们仍以纤维素和淀粉为例分别介绍前两类结构的特点。

11.3.1 纤维素的立体结构

纤维素存在于大多数植物细胞壁中。纤维素分子中，β-D-葡萄糖结构单元以β-1,4-糖苷酯键连接成链形构象的多聚物，见图11.1。纤维素的分子量高达5万以上。约2000条链状的纤维素分子以氢键及其他非共价结合的相互作用形成捆，直径约10—25nm；每2000个左右的捆构成一根微纤维(microfibril)；由许多微纤维再组成在光学显微镜下可以看到的一个粗纤维(macrofibril)。由X射线纤维衍射花样所计算和推测的天然苎麻纤维素的模型见图11.8[130]。这是纤维素Ⅰ的一个反平行排列的模型。链状的天然纤维素分子反平行交错排列，并以氢键相联系堆积成片层状的三维结构。

11.3.2 淀粉的立体结构

淀粉几乎与纤维素一样广泛地存在于植物中。它与动物的糖元相似，是贮存能量的物质。淀粉分子在淀粉粒中堆积贮藏起来。这些贮备物的堆积是结构化的，因而在一般情况下不被干扰，而一旦需要时，淀粉粒中的淀粉分子极易受到影响而被充分地利用。直链淀粉是α-D-葡萄糖结构单元以α-1,4-糖苷酯键所连接成的多聚物，分子量约为6万左右，相当于300—400个葡萄糖分子缩合而成。

尽管关于淀粉的X射线衍射花样的首次报道是早在20世纪30年代[128]，但是要对其结晶性排列得到毫不含糊的了解则至今尚未完全成功。人们已提出了若干种结构模型，其中K. Kainuma和D. French[131]于1972年提出的B-多晶型淀粉模型为一个双股的螺旋。而现在被人们所接受的模型是H-C. H. Wu等人[132,133]从X射线衍射研究所推测出的A型直链淀粉和B

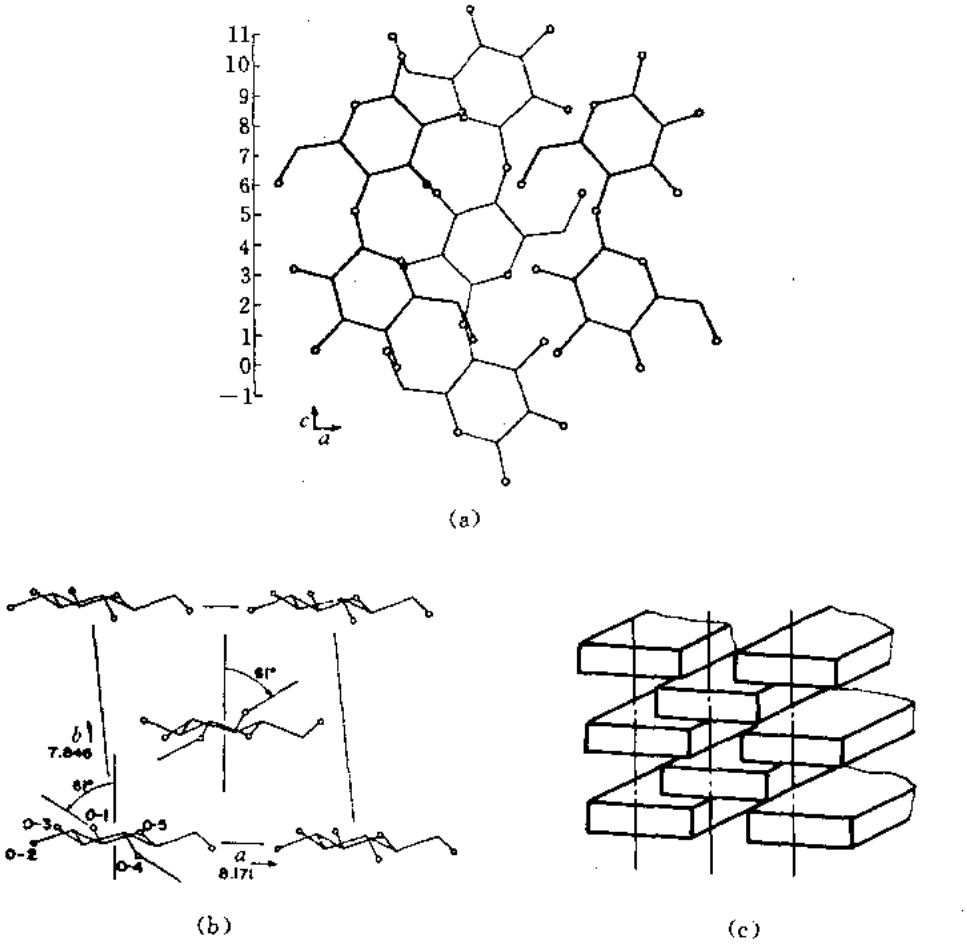


图11.8　天然苎麻纤维素的结构和堆积方式的模型

(a)垂直于纤维轴方向的投影;(b)沿纤维轴方向的投影;(c)纤维素分子的堆积方式示意图

型直链淀粉的模型,分别示于图11.9和图11.10。两者均为反平行堆积的右手双股螺旋模型,是具有空心螺旋的类型。每股螺旋中每圈为6个葡萄糖残基,即螺旋分子具有六重螺旋轴对称性,重复周期为2.08nm。但是,两个模型中的双螺旋在晶胞中的堆积方式有相当大的差别。在B型直链淀粉模型中,双螺旋围绕着一个大的空腔排列着,空腔中有水分子;而在A型直链淀粉模型中,此空腔为一个双股螺旋所占据。

必须指出的是,上面所描述的纤维素和淀粉的分子结构以及它们的堆积方式等,仍然只是到现在为止较为合理的模型而已。关于它们的结构仍在继续进行研究,而且除了X射线衍射分析法以外,还报道了用电子衍射方法、中子衍射方法以及核磁共振方法等的研究结果,对一些问题还有争议。例如淀粉结构中究竟是左手还是右手螺旋?每圈螺旋有几个葡萄糖残基?是单股螺旋还是双股螺旋?螺旋之间堆积的方式是平行的还是反平行的等等。例如对于上述现在为人们所接受的A型淀粉的模型已有不少学者提出异议,1988年A. Imberty 等人[134]提出了一个与其大不相同的新模型,见图11.11和图11.12。在这个模型中,晶格为单斜点阵,$a=2.124\text{nm}$,$b=1.172\text{nm}$,$c=1.069\text{nm}$,$\gamma=123.5°$;空间群为B2;每个晶胞中含有12个葡萄糖残基,它们位于两个左手平行的双螺旋中,这两个双螺旋以平行的方式堆积着,螺旋轴平行于c

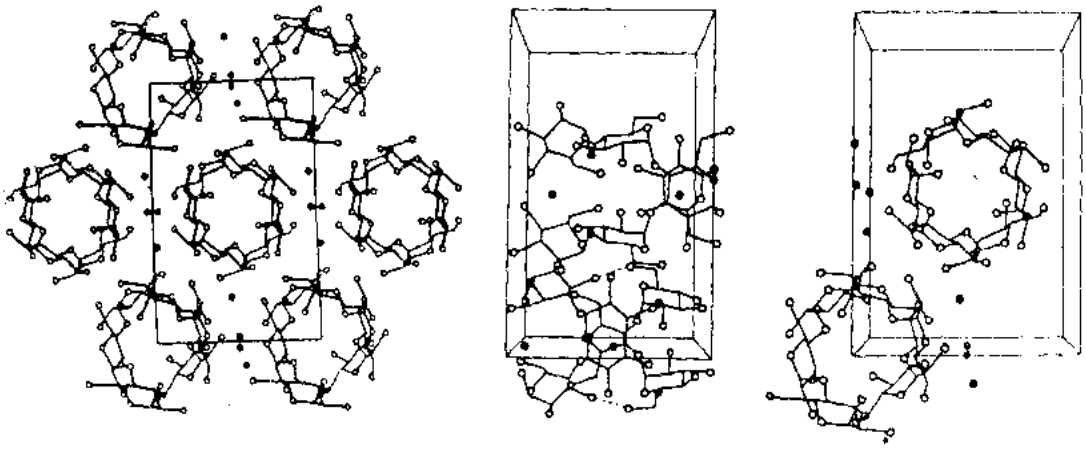


图11.9　A型直链淀粉的分子结构和堆积方式的模型

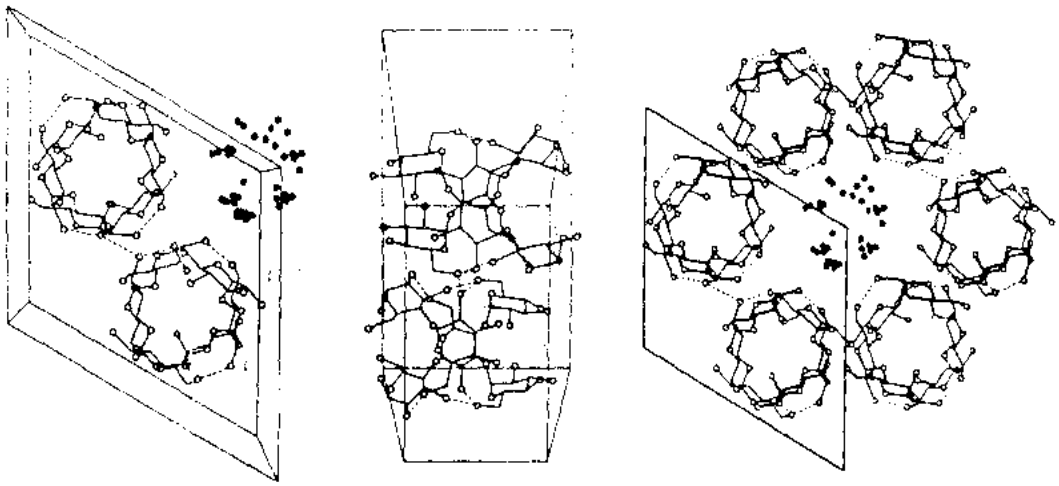


图11.10　B型直链淀粉的分子结构和堆积方式的模型

轴;4个水分子位于这些螺旋之间。这个新的结构模型是在综合了A型淀粉的单晶电子衍射、X射线粉末衍射花样以及根据过去已报道过的X射线纤维衍射数据重新指标化后所分析的结果而提出来的。

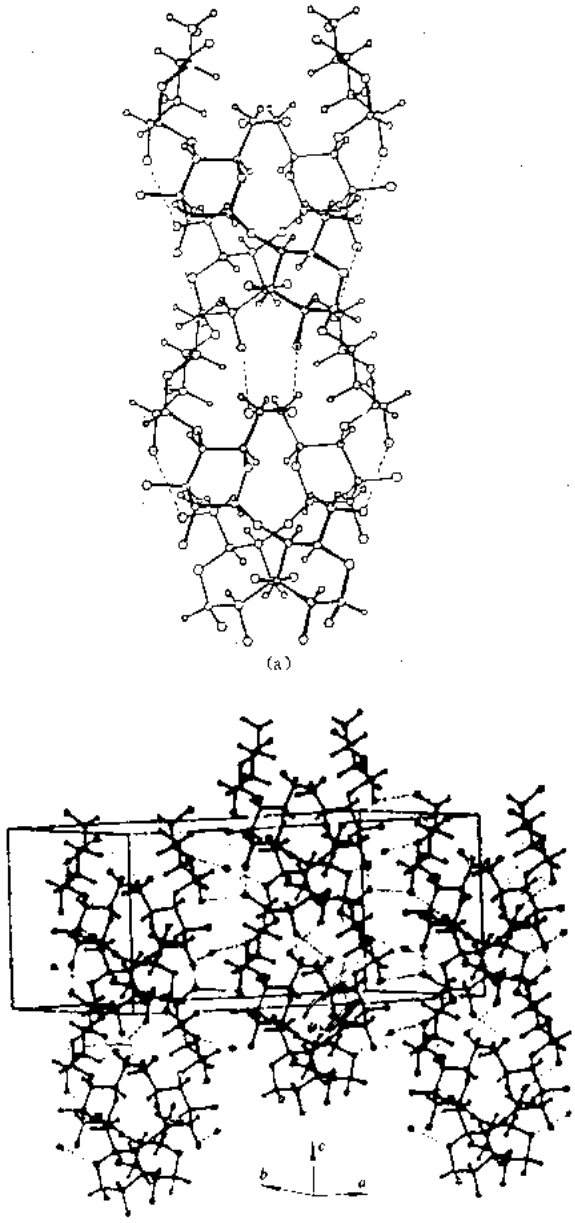


图11.11　A 型淀粉的左手平行双螺旋模型

(a)两条左手螺旋的直链淀粉单链所形成的平行左手双螺旋，每条链是以$2c=2.138nm$为螺距的六重螺旋重复，虚线代表氢键；(b)每个晶胞中的两个左手平行双螺旋以平行的方式堆积着，(●)代表水分子，虚线代表氢键

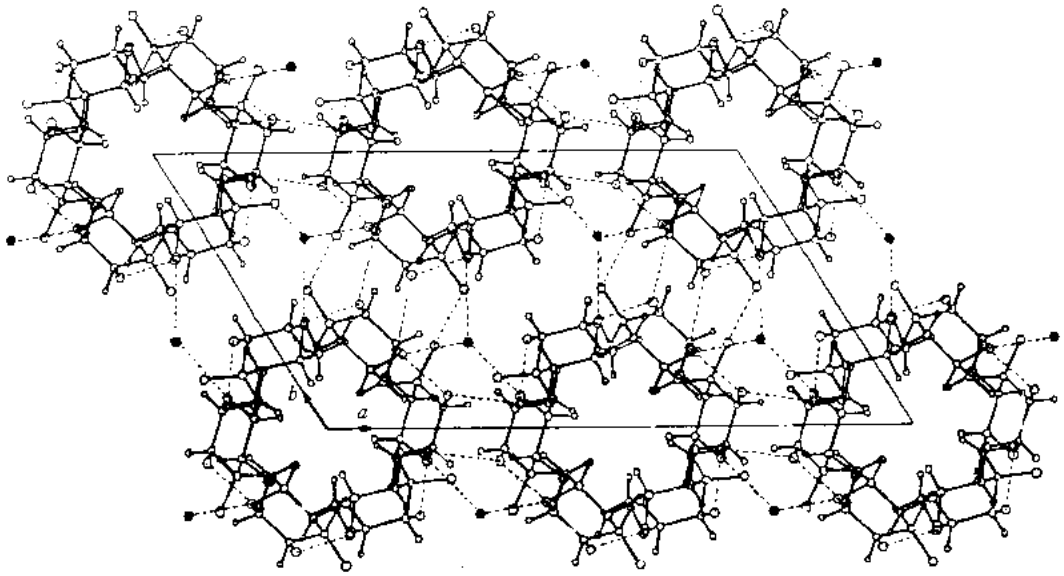


图11.12　A 型淀粉的结构在(a,b)平面上的投影
(●)代表水分子;虚线代表氢键

参 考 文 献

[1] 唐有祺，结晶化学，高等教育出版社，1957.

[2] 周公度，晶体结构测定，科学出版社，1981.
周公度，结构化学基础，北京大学出版社，1989.

[3] 张绍辉、傅享，X 射线晶体结构分析导论，武汉大学出版社，1989.

[4] 梁栋材，X 射线晶体学基础，科学出版社，1991.

[5] G. H. W. Milburm 著，王守道译，X 射线晶体学，人民出版社，1984.

[6] Arne Engstom & J. B. Finean 著，叶容译，生物超微结构，人民出版社，1979.

[7] 谢有畅、邵美成，结构化学，人民出版社，1979.

[8] 北京医学院生物物理教研组，生物物理学，人民卫生出版社，1961.

[9] 陶慰孙等，蛋白质分子基础，人民教育出版社，1981.

[10] T. L. Blundell & L. N. Johnson, Protein Crystallography, Academic Press Inc. New York and London, 1976.

[11] B. S. Richard, Molecular Biophysics, Addison-Wesley Publishing Company, Inc., 1962.

[12] M. J. Buerger, "X-Ray Crystallography, John Wiley, 1942.

[13] C. W. Bunn, Chemical Crystallography, Oxford University Press, 1945.

[14] H. S. George, X-ray Structure Determination, Macrmillan Publishing Company, Inc., 1968.

[15] H. W. Wyckoff, Methods in Enzymology, Volume 114, 1985.

[16] M. J. Buerger, The Precession Method, John Wiley, 1964.

[17] S. C. Burg, X-Ray Analysis of Organic Structures, Academic Press Inc. New York and London, 1961.

[18] H. Tadokoro, Structure of Crystalline Polymers, Wiley-interscience Publication, 1979.

[19] G. Zubay, Biochemistry, Addison-Wesley Publishing Company, Inc., 1983.

[20] IUPAC-IUB Commission on Biochemical Nomenclature, *J. Mol. Biol.* (1970), **52**, 1.

[21] G. N. Ramachandran, C. Ramakrishnan & V. Sasisekharan, *J. Mol. Biol.* (1963), **7**, 95.

[22] G. N. Ramachandran & V. Sasisekharan, *Advanc. Protein Chem.* (1968), **23**, 283.

[23] W. T. Astbury & A. Street, *Phil. Trans. Roy. Soc.* (London), (1931), A**230**, 75.

[24] L. Pauling & R. B. Corey, *Proc. Natl. Acad. Sci.* (1951), **37**, 205.

[25] L. Pauling & R. b. Corey, *Proc. Natl. Acad. Sci.* (1951), **37**, 251.

[26] J. B. Sumner, *J. Biol. Chem.* (1926), **69**, 435.

[27] J. D. Bernal & D. C. Crowfoot, *Nature* (1934), **133**, 794.

[28] 卢光莹、毛春虹、卫新成、华子千、顾孝诚，生物物理学报(1989)，**5**(3)，259.

[29] 华子千、朱红、卫新成、卢光莹、顾孝诚，生物物理学报(1990)，**6**(2)，221.

[30] 卢光莹、华子千、周岚、金雷、卫新成、郭虹、倪逸声、张龙翔、李根培、唐有祺、戚正武，生物化学杂志(1989)，**5**(1)，24.

[31] 卢光莹、温晓刚、颜岩、李松林、华子千、顾孝诚、刘进元、陈章良，生物物理学报(1994)，**10**(2)，193.

[32] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips & V. C. Shore, *Nature*(1960), **185**, 442.

[33] D. W. Green, V. M. Ingram & M. F. Perutz, *Proc. Roy, Soc.* (1954), A**225**, 287.

[34] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wykoff & D. C. Phillips, *Nature*(1958), **181**, 622.

[35] J. C. Spurlino, Guang ying Lu & F. A. Quiocho, *J. Biol. Chem.* (1991), **266**(8), 5202.
[36] C. C. F. Blake, D. F. Koening, G. A. Mair, A. C. T. North, D. C. Phillips & V. R. Sarma, *Nature* (1965), **206**, 757.
[37] G. Kartha, J. Bello & D. Harker, *Nature* (1967), **213**, 862.
[38] H. W. Wyckoff, K. D. Hardman, N. M. Allewell, T. Inagami, D. Tsernoglou, L. N. Johnson & F. M. Richards, *J. Biol. Chem.* (1967), **2427**, 3984.
[39] B. W. Matthews, P. B. Sigler, R. Henderson & D. M. Blow, *Nature* (1967), **214**, 652.
[40] G. N. Reeke, J. A. Hartsuck, M. L. Ludwig, F. A. Quiocho, T. A. Steitz & W. N. Lipscomb, *Proc. Natl. Acad. Sci.* USA (1967), **58**, 220.
[41] J. Deisenhofer, O. Epp, K. Miki, R. Huber & H. Michel, *J. Mol. Biol.* (1984), **180**, 385.
[42] 胰岛素结构研究组, 中国科学(1972), **1**, 1.
[43] J. S. Richardson, in Advance in Protein Chemistry (1981), Vol. 34, pp. 167—339, Edited by C. B. Anfinsen, J. T. Edsall & F. M. Richards, Academic Press, New York and London.
[44] B. K. Vainshtein, V. M. Fridkin & V. L. Indenbom, Modern Crystallography II——Structure of Crystals (1982), Springer Verlag, Berlin, Heidelberg and New York.
[45] J. S. Richardson & D. C. Richardson, in Prediction of Protein Structure and The Principles of Protein Conformation (1989), pp. 1—98, Edited by G. D. Fasman, Plenum Press, New York and London.
[46] J. F. Leszczynski & G. D. Rose, Science (1986), **234**, 849.
[47] J. S. Richardson, *Proc. Natl. Acad. Sci.* USA (1976), **73**, 2619.
[48] J. S. Richardson, *Nature* (1977), **268**, 495.
[49] M. Levitt & C. Chothia, *Nature* (1976), **261**, 552.
[50] D. B. Wetlaufer, *Proc. Natl. Acad. Sci.* USA (1973), **70**, 697.
[51] G. Rose, *J. Mol. Biol.* (1979), **134**, 447.
[52] J. I. Harris & C. A. Knight, *J. Biol. Chem.* (1955), **214**, 215.
[53] H. T. Wright, Hua Zi-qian & R. Hüber, *J. Mol. Biol.* (1990), **213**, 513.
[54] M. F. Perutz, H. Muirhead, J. M. Cox & L. C. G. Goaman, *Nature* (1968), **219** (5150), 131.
[55] W. Bolton & M. F. Perutz, *Nature* (1970), **228** (5271), 551.
[56] R. C. Ladner, E. J. Heidner & M. F. Perutz, *J. Mol. Biol.* (1977), **114**(3), 385.
[57] P. W. Tucker & M. F. Perutz, *J. Mol. Biol.* (1977), **114**(3), 415.
[58] Crystallization of Membrane Protein, Edited by H. Michel, CRC Press, Boca Raton, Ann Arbor, Boston, 1990.
[59] H. Michel, *Trends in Biochemical Sciences* (1983), **8**, 56.
[60] H. Michel & D. Oesterhelt, Proc. *Natl. Acad. Sci.* USA (1980), **77**, 1283.
[61] H. Michel, *J. Mol. Biol.* (1982), **158**, 567.
[62] J. Deisenhofer, O. Epp, K. Miki, R. Hüber & H. Michel, *Nature* (1985), **318**, 618.
[63] M. Garavito & J. P. Rosenbush, *J. Cell Biol.* (1980), **86**, 327.
[64] J. P. Allen, G. Feher, T. O. Yeats, D. C. Rees, D. C. Eisenberg, J. Deisenhofer, H. Michel & R. Hüber, *Biophys. J.* (1986), **49**, 583.
[65] B. Kleffel, R. M. Garavito, W. Baumeister & J. P. Rosenbusch, *EMBO J.* (1985), **4**(6).
[66] M. Tomita & V. T. Marchesi, *Proc. Natl. Acad. Sci.* USA (1975), **72**, 2964.
[67] I. A. Wilson, J. J. Skehel & D. C. Wiley, *Nature* (1981), **289**, 366.
[68] D. W. Urry, *Proc. Natl. Acad. Sci.* USA (1971), **68**, 672.
[69] H. Vogel & F. Jähnig, *Biophys. J.* (1986), **50**, 573.

[70] R. O. Fox & F. M. Richards, *Nature* (1982), **300**, 325.
[71] R. Henderson & P. N. T. Unwin, *Nature* (1975), **257**, 28.
[72] H. Vogel, J. K. Wright & F. Jähnig, *EMBO J.* (1985), **4**, 3625.
[73] H. Vogel & F. Jähnig, *J. Mol. Biol.* (1986), **190**, 191.
[74] W. K. Olson & P. J. Flory, *Biopolymers* (1972), **11**, 1.
[75] J. D. Watson & F. H. C. Crick, *Nature* (1953), **171**, 737.
[76] W. T. Astbury & F. O. Bell, *Nature* (1938), **141**, 747.
[77] R. E. Franklin & R. G. Gosling, *Nature* (1953), **171**, 740.
[78] R. Langridge, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins & L. D. Hamilton, *J. Biophys. Biochem. Cytol.* (1957), **3**, 767.
[79] M. H. F. Wilkins & S. Arnott, *J. Mol. Biol.* (1965), **11**, 391.
[80] M. A. Mokul'sky, K. A. Kapitonova & T. D. Mokul'skaya, *Mol. Biol.* USSR (1972), **6**, 716.
[81] M. A. Viswamitra, O. Kennard, P. G. Jones, G. M. Sheldrick, S. Salisbury, L. Falvello & Z. Shakked, *Nature* (1978), **273**, 687.
[82] A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Mare & A. Rich, *Nature* (1979), **282**, 680.
[83] R. M. Wing, H. R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura & R. E. Dickerson, *Nature* (1980), **287**, 755.
[84] Biological Macromolecules & Assemblies, Volume 2: Nucleic Acids & Interactive Proteins, Edited by F. A. Jurnak and A. McPherson, John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1985.
[85] S. H. Kim, G. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl & A. Rich, *J. Mol. Biol.* (1973), **75**, 421.
[86] S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussmann, A. H. J. Wang, N. C. Seeman & A. Rich, *Science* (1974), **185**, 435.
[87] J. D. Robertus, Y. E. Ladner, Y. T. Finch, D. Rhodes, R. S. Brown, B. F. C. Clark & A. Klug, *Nature* (1974), **250**, 546.
[88] F. H. C. Crick & J. D. Watson, Virus Structure: General Principles, in The Nature of Viruses, Edited by G. E. W. Wolstenholme & E. C. P. Millar, pp. 5—13, J. & R. Churchill Ltd., London, 1957.
[89] W. M. Stanley, *Science* (1935), **81**, 644.
[90] J. D. Bernal, I. Fankuchen & D. P. Riley, *Nature* (1938), **142**, 1075.
[91] J. T. Finch & A. Klug, *Nature* (1959), **183**, 1709.
[92] F. C. Bawden, N. W. Pirie, J. D. Bernal & I. Fankuchen, *Nature* (1936), **138**, 1051.
[93] J. D. Bernal & I. Fankuchen, *J. Gen. Physiol.* (1941), **25**, 111.
[94] R. W. G. Wyckoff & R. B. Corey, *J. Biol. Chem.* (1936), **116**, 51.
[95] G. Stubbs, S. Warren & K. Holmes, *Nature* (1977), **267**, 216.
[96] S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler & G. Bricogne, *Nature* (1978), **276**, 368.
[97] J. Gregory & K. C. Holmes, *J. Mol. Biol.* (1965), 796.
[98] R. Franklin & K. Holmes, *Acta Crystallogr.* (1958), **11**, 213.
[99] L. Makowski, *J. Appl. Crystallogr.* (1982), **15**, 546.
[100] G. Stubbs, K. Namba & L. Makowski, *Biophys. J.* (1986), **49**, 58.
[101] R. Leberman, J. T. Finch, P. F. C. Gilbert, J. Witz & A. Klug, *J. Mol. Biol.* (1974), **86**, 179.
[102] A. C. Bloomer, J. N. Champness, G. Bricogne, R. Staden & A. Klug, *Nature* (1978), **276**, 362.

[103] D. L. D. Caspar & A. Klug, *Cold Spring Harbor Symp.* (1962), **27**, 1.
[104] L. Makowski, Structure Diversity in filamentous Bacteriophages, in Biological Macromolecules and Assemblies, Vol. I; Virus Structures, Edited by F. A. Jurnak & A. McPherson, pp. 203—253, Wiley, New York, 1984.
[105] L. Makowski, D. L. D. Caspar & D. A. Marvin, *J. Mol. Biol.* (1980), **140**, 149.
[106] C. Nave, R. S. Brown, A. G. Fowler, J. E. Ladner, D. A. Marvin, S. W. Provencher, A. Tsugita, J. Armstrong & R. N. Perham, *J. Mol. Biol.* (1981), **149**, 675.
[107] D. L. D. Caspar & A. Klug, *Cold Spring Harbor Symp.* (1962), **27**, 1.
[108] T. A. Jones & L. Liljas, *J. Mol. Biol.* (1984), **177**, 735.
[109] C. Abad Zapatero, S. S. Abdel-Meguid, J. E. Johnson, A. G. W. Leslie, I. Rayment, M. G. Rossmann, D. Suck & T. Tsukihara, *Nature* (1980), **286**, 33.
[110] T. J. Smith, M. J. Kremer, M. Luo, G. Vriend, E. Arnold, G. Kamer, M. G. Rossmann, M. A. McKinlay, G. D. Diana & M. J. Otto, *Science* (1986), **233**, 1286.
[111] J. M. Hogle, M. Chow & D. J. Filman, *Science* (1985), **229**, 1358.
[112] M. Luo, G. Vriend, G. Kamer, I. Minor, E. Arnold, M. G. Rossmann, U. Boege, D. G. Scraba, G. M. Duke & A. C. Palmenberg, *Science* (1987), **235**, 182.
[113] M. M. Roberts, J. L. White, M. G. Grutter, R. M. Burnett, *Science* (1986), **232**, 1148.
[114] F. K. Winkler, C. E. Schutt, S. C. Harrison & G. Bricogne, *Nature* (1977), **265**, 509.
[115] G. Stubbs, Virus Structure, in Prediction of Protein Structure and the Principles of Protein Conformation, Edited by G. D. Fasman, Plenum Press, New York and London, 1989.
[116] F. A. Jurnak & A. McPherson eds., Biological Macromolecules and Assembles, Vol. I; Virus Structure, Wiley, New York, 1984.
[117] J. P. Langmore & J. R. Paulson, *J. Cell Biol.* (1983), **96**, 1120.
[118] J. R. Paulson & J. P. Langmore, *J. Cell Biol.* (1983), **96**, 1132.
[119] J. T. Finch & A. Klug, *Proc. Natl. Acad. Sci.* USA (1976), **73**(6), 1897.
[120] L. Sperling & A. Klug, *J. Mol. Biol.* (1977), **112**, 253.
[121] F. Thoma, T. H. Koller & A. Klug, *J. Cell Biol.* (1979), **83**, 403.
[122] J. T. Finch, L. C. Lutter, D. Rhodes, R. S. Brown, B. Rushton, M. Levitt & A. Klug, *Nature* (1977), **269**, 29.
[123] J. T. Finch, R. S. Brown, D. Rhodes, T. Richmond, B. Rushton, L. C. Lutter & A. Klug, *J. Mol. Biol.* (1981), **145**, 757.
[124] T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes & A. Klug, *Nature* (1984), **311**, 532.
[125] A. Klug, D. Rhodes, J. Smith, J. T. Finch & J. O. Thomas, *Nature* (1980), **287**, 509.
[126] R. W. Burlingame, W. E. Love, B-C. Wang, R. Hamlin, N-H. Xuang & E. N. Moudrianakis, *Science* (1985), **228**, 546.
[127] C. Trogus & K. Hess, *Z. Physik. Chem. ABT.* (B)(1930), **6**, 1.
[128] J. R. Katz & T. B. van Italie, *Z. Physik. Chem.* (A) (1930), **150**, 90.
[129] H-C. H. Wu & A. Sarko, *Carbohyd. Res.* (1977), **54**(1), c3-c6.
[130] A. D. Frence, Carbohyd. Res. (1978), **61**, 67.
[131] K. Kainums & D. French, *Biopolymers* (1972), **11**, 2241.
[132] H-C. H. Wu & A. Sarko, *Carbohyd. Res.* (1978), **61**, 7.
[133] H-C. H. Wu & A. Sarko, *Carbohyd. Res.* (1978), **61**, 27.
[134] A. Imberty, H. Chanzy, S. Pérez, A. Buléon & V. Tran, *J. Mol. Biol.* (1988), **201**, 365.